Randomness-Efficient Rumor Spreading

Zeyu Guo*
California Institute of Technology
Pasadena, USA
zguo@caltech.edu

He Sun†
Max Planck Institute for Informatics
Saarbrücken, Germany
hsun@mpi-inf.mpg.de

Abstract

We study the classical rumor spreading problem, which is used to spread information in an unknown network with n nodes. We present the first protocol for any expander graph G with n nodes and minimum degree $\Theta(n)$ such that, the protocol informs every node in $O(\log n)$ rounds with high probability, and uses $O(\log n \log\log n)$ random bits in total. The runtime of our protocol is tight, and the randomness requirement of $O(\log n \log\log n)$ random bits almost matches the lower bound of $\Omega(\log n)$ random bits. We further study rumor spreading protocols for more general graphs, and for several graph topologies our protocols are as fast as the classical protocol and use $\widetilde{O}(\log n)$ random bits in total, in contrast to $O(n \log^2 n)$ random bits used in the well-known rumor spreading push protocol. These results together give us almost full understanding of the randomness requirement for this basic epidemic process.

Our protocols rely on a novel reduction between rumor spreading processes and branching programs, and this reduction provides a general framework to derandomize these complex and distributed epidemic processes. Interestingly, one cannot simply apply PRGs for branching programs as rumor spreading process is not characterized by small-space computation. Our protocols require the composition of several pseudorandom objects, e.g. pseudorandom generators, and pairwise independent generators. Besides designing rumor spreading protocols, the techniques developed here may have applications in studying the randomness complexity of distributed algorithms.

Keywords: distributed computing, rumor spreading, randomness complexity, branching programs, pseudorandomness

1 Introduction

Rumor spreading is one of the most important communication primitives in large networks, and has been studied under different names such as gossip, information dissemination, or broadcasting. Efficient protocols for information spreading have wide applications in failure detection [35], resource discovery [28], replicated database systems [11, 18], and modeling the spread of computer viruses [4]. Besides computer science, the dynamics of such processes in social networks also constitutes a research topic in economics and sociology.

The simplest and widely studied form of information spreading protocols is the so-called *push model* of rumor spreading. Initially, a message, called *a rumor*, is placed on an arbitrary node of an unknown network with n nodes. In subsequent synchronous rounds, every node that knows the rumor picks a neighbor uniformly at random and sends the rumor to the chosen neighbor. This process continues until every node gets the rumor. It was shown that this simple protocol is very efficient for several network topologies [15, 16, 18, 21]. In particular, its

*This work is supported by NSF grant CCF-1116111. Part of this work was done while visiting Max Planck Institute for Informatics.

†Part of this work was done while visiting California Institute of Technology.

runtime, the number of rounds required until every node gets the rumor with high probability, is logarithmic in the number of nodes in the graph. Graphs satisfying this property range from complete graphs, hypercubes, Erdös-Réyi random graphs, and "quasi-regular" expanders (i.e., expander graphs for which the ratio between the maximum and minimum degree is constant). Besides the efficiency, this protocol is local (i.e. no knowledge of global graph structure is needed), simple (e.g. nodes do not have initial IDs, in contrast to the protocol in [27]), and can tolerate the failure of some links. More recently, several variations of information spreading protocols were proposed to allow information to spread efficiently on *arbitrary* networks [8], or networks with weak expansion properties [7].

Most of these algorithms are inherently randomized in both their design and analysis in that they crucially rely on the effect of choosing neighbors randomly for every node among different rounds. However, it is not clear if this randomization is essential for efficiently disseminating the information. Hence the randomness requirement, the number of random bits used in total in order to spread the rumor efficiently, becomes a key measurement for designing rumor spreading protocols. One of the most studied questions concerns the randomness requirement: how many random bits are sufficient to efficiently spread a rumor to all nodes in a graph? While for any graph with n nodes, the above-mentioned *fully-random* push protocol that finishes in T rounds needs $O(T \cdot n \log n)$ random bits, it is not difficult to show that for any graph G of n nodes, there is a protocol which uses $3 \log n$ random bits in total, and whose runtime is as fast as the standard fully-random protocol (See Corollary B.2 for the formal statement). However, the explicit construction of such protocols is much more difficult, and a long line of research has been devoted to finding randomness-efficient protocols, see [2, 12–14, 23, 24, 27] for instance. Prior to work, the best results need to use $\Omega(\log^2 n)$ random bits, as the randomness used by different rounds have to be independent, and $\Omega(\log n)$ random bits are necessary per round.

Our Results. The present paper constructs randomness-efficient rumor spreading protocols. Our main results are as follows:

- Let G be an expander graph with n nodes and minimum degree $\delta = \Theta(n)$. Then there is a protocol using $O(\log n \log\log n)$ random bits in total, such that every node gets informed in $O(\log n)$ rounds with high probability (cf. Corollary 3.4).
- Let G be a general graph with n nodes, conductance ϕ and the ratio between maximum and minimum degree $\Delta/\delta = O(1)$. Then there is a protocol using $O((1/\phi)\log n(\log\log n + \log\Delta))$ random bits in total, such that every node gets informed in $O((1/\phi)\log n)$ rounds with high probability (cf. Theorem 4.1).

Note that any protocol needs at least $\Omega(\log n)$ rounds to spread the rumor to all nodes, hence the runtime $O(\log n)$ rounds in our first result is tight. For the randomness requirement, our first protocol needs $O(\log n \log\log n)$ random bits, which improves the previous best bound of $O(\log^2 n)$ random bits in total [24]. Since for any expander graph with n nodes and minimum degree $\delta = \Theta(n)$, any protocol that finishes in $O(\log n)$ rounds with high probability needs at least $\Omega(\log n)$ random bits (cf. Theorem C.2), our bound is almost tight.

The second result is for general graphs. The runtime here matches the upper bound known in the truly random protocol, and is tight, in the sense that there are graphs with diameter $\Omega((1/\phi)\log n)$ [9]. The randomness complexity of the second result improves the previous best one of $O((1/\phi)\log^2 n)$ random bits. See Table 1 for a comparison between our results and previous best ones.

Techniques. Our main result is based on a generic reduction between the problem of designing rumor spreading protocols and the problem of constructing pseudorandom generators (PRGs)

Graph family	Rumor spreading time	Random bits	Reference
Expander Graphs $\delta = \Theta(n)$	$O(\log n)$,	$\Omega(\log n)$ [Lower Bound]	Theorem C.3
		$O(\log^2 n)$ [Previous Best Result]	[22]
		$O(\log n \log\log n)$	Theorem 3.3
Graphs with $\Delta/\delta = O(1)$	$O((1/\phi)\log n)$,	$\Omega(n \log n \log \Delta)$	[21, 31]
		$O((1/\phi)\log^2 n)$ [Previous Best Result]	[24]
		$O((1/\phi)\log n \cdot (\log\log n + \log \Delta))$	Theorem 4.1
Strong Expanders	$\log n + \ln n + o(\log n)$,	$\Theta(n \log n \log \Delta)$	[24]
		$O(\log^3 n)$ [Previous Best Result]	[24]
		$O(\log n \cdot (\log\log n + \log \Delta))$	Theorem 5.1
$G(n,p)$ with $p = \omega(\log n/n)$	$\log n + \ln n + o(\log n)$,	$\Theta(n \log n \log(pn))$	[24] & [20]
		$O(\log^3 n)$ [Previous Best Result]	[24]
		$O(\log n \cdot (\log\log n + \log \Delta))$	Theorem 5.1

Table 1: Comparison of the time needed to spread a rumor and the required number of random bits for various topologies. By Δ we denote the maximum degrees of G, and by ϕ the conductance of G. See Section 3 for the formal definition of conductance ϕ.

for branching programs. This key insides is new in the area of studying epidemic algorithms. In retrospect, this connection between the two problems is natural because (1) random walks over branching programs resemble the rumor spreading process where nodes send messages to random neighbors, and (2) in a rumor spreading protocol, each node has access to only its own neighboring list, and is oblivious to the structure of the network. This is an analogue of *oblivious derandomization* achieved by PRGs.

However, rumor spreading appears much more complicated than small-space computation due to the following facts: (1) In the rumor spreading process, rumors are "duplicated" every round, although every "existing" rumor viewed individually performs a random walk. Hence instead of considering every single random walk performed by any fixed rumor, we need to study the dynamics behind the whole rumor spreading process. (2) The state of the process at some time essentially depends on the past behavior of all nodes and is by no means computable in small space. Indeed, even knowing if a single node u gets the rumor at some round requires knowing the set of its neighbors having the rumor in the previous rounds, and may require $\deg(u) = \Theta(n)$ bits for dense graphs. For these reasons, this connection to small-space computation is delicate and not obvious.

Surprisingly, we show that a reduction exists from the problem of designing randomness-efficient rumor spreading protocols to the problem of constructing PRGs for small-space computation. Hence the question of designing randomness-efficient rumor spreading protocols is now exposed to the numerous techniques used in the study of constructing PRGs for small-space computation. In particular, the explicit constructions (resp. existence) of PRGs fooling certain branching programs imply the explicit constructions (resp. the existence) of rumor spreading protocols, and the explicit constructions of PRGs with optimal parameters imply randomness-optimal rumor spreading protocols for dense graphs.

For general graphs, our present a protocol that uses independent seeds of length $O(\log\log n + \log \Delta)$ in each round, where Δ is the maximum degree of the graph. The key idea is that the rumor spreading process enjoys nice locality when the maximal degree is small. Our protocol take advantage of this feature by using an class of objects called unbalanced expanders, which are then composed with pairwise independent generators. This technique yields much smaller seed length than using pairwise independent generators alone. This protocol has the advantage of being very simple. Furthermore, a simple variant of this protocol using PRGs for combinatorial rectangles achieves the tight runtime for strong expanders.

Related Work. There is a large amount of literature devoted to various aspects of rumor spreading. The majority of research studies the rumor spreading time in terms of the graph

properties, such as conductance [21], vertex expansion [22], mixing time [5], diameter [18] and degree [18]. For instance, the first explicit connection between randomized rumor spreading and graph expansion was established by Mosk-Aoyama and Shah [31], who proved that on any regular graph with conductance ϕ, the protocol finishes in $O((1/\phi) \cdot \log n)$ rounds.

The study of randomness complexity of rumor spreading protocols was started by Doerr et al. [13], who proposed a *quasi-random* version of the above-mentioned push protocol. This quasi-random protocol is as follows: Every node u has a (cyclic) list of its neighbors. Once node u is informed, it starts at a random position of the list, but from then on node u informs its neighbors in the order of the list. In contrast to $O(n \log^2 n)$ random bits that used in the standard push model, they show that, by using $O(n \log n)$ random bits, the quasi-random protocol is as fast as the *truly random* protocol for many graph topologies, e.g. complete graphs [2, 19], random graphs, hypercubes or expanders [14]. It is known that for this quasi-random model one cannot further reduce this amount without a severe loss of efficiency [12]. Giakkoupis and Woelfel [23] derived a protocol, which used $O(n \log \log n)$ random bits in total and finishes in $O(\log n)$ rounds on complete graphs of n nodes. Recently, Giakkoupis et al. [24] presented two protocols. In contrast to previous work, the protocols in [24] use only a poly-logarithmic number of random bits, and are as fast as the truly random protocol. Besides the rumor spreading problem, researchers also studied the question of designing randomness-efficient or deterministic protocols for similar problems. For instance, Haeupler [27] presented one deterministic gossip algorithm for the local broadcast problem.

Throughout the paper we assume that nodes have no initial IDs, and we combine the protocols with an ID distribution mechanism so that every node gets a unique ID once it gets the rumor. Moreover, we assume the *standard adversary model*, which was also used in [13, 23, 24]: In each round, every informed node u chooses an index $j \in \{1, \ldots, \deg(u)\}$, and sends a rumor to the jth node in its adjacency list. No edge connection information is available to u other than its adjacency list; and the order of u's neighbors in this list is determined by an oblivious adversary (before the algorithm is executed).

Organization. The paper is structured as follows: We review some basic notations and tools that we will use in Section 2. Section 3 presents the protocols for dense expander graphs, and the reductions between the problem of rumor spreading and the problem of constructing PRGs for branching programs. In Section 4 we construct one protocol for general graphs. The protocol for strong expander graphs will be discussed in Section 5.

2 Preliminaries

Let $G = (V, E)$ be a connected, and undirected graph with n nodes. The maximum, minimum, and average degree of G are represented by Δ, δ, and d. For any node u, the degree of u is represented by $\deg(u)$, and the set of neighbors of u is represented by $N(u)$. Moreover, for any set $S \subseteq V$, the neighboring set of S is defined by $N(S) \triangleq \bigcup_{u \in S} N(u)$, and the volume of S is defined by $\mathrm{vol}(S) \triangleq \sum_{u \in S} \deg(u)$. For any set $S, T \subseteq V$, we define $E(S, T) \triangleq \{\{u, v\} : u \in S \text{ and } v \in T\}$ and $e(S, T) \triangleq |E(S, T)|$.

By $\log x$ we denote the binary logarithm of x. For any integer m, define $[m] \triangleq \{0, \ldots, m-1\}$. The product distribution of two distributions X and Y is denoted by $X \times Y$. The disjoint union of a family of sets $\{A_i : i \in I\}$ indexed by I is denoted by $\bigsqcup_{i \in I} A_i \triangleq \bigcup_{i \in I} \{(x, i) : x \in A_i\}$. With high probability stands for with probability $1 - o(1)$.

We introduce the main tools that we use in our paper.

2.1 Pseudorandom Generators

One key tool used in our construction is pseudorandom generators. Informally, pseudorandom generators are deterministic algorithms f such that, for a given random string x chosen uniformly at random from set $\{0,1\}^\ell$, f produces a string $f(x) \in \{0,1\}^{n(\ell)}, n(\ell) \geqslant \ell$, which is indistinguishable from the uniform distribution over $\{0,1\}^{n(\ell)}$ by a certain family $\mathcal{C}$ of functions.

Pairwise Independent Generators.

Definition 2.1 (Pairwise Independent Generator)**.** *We say $X_0, \ldots, X_{d-1}$ with X_i distributed over $[m_i]$ are ε-pairwise independent if*

- $\left|\mathbf{Pr}\left[\, X_i = x \,\right] - \frac{1}{m_i}\right| \leqslant \varepsilon$ *for all $i \in [d]$ and $x \in [m_i]$, and*
- $\left|\mathbf{Pr}\left[\, X_i = x \wedge X_j = x' \,\right] - \frac{1}{m_i \cdot m_j}\right| \leqslant \varepsilon$ *for all distinct $i, j \in [d]$ and all $x \in [m_i]$, $x' \in [m_j]$.*

We say they are pairwise independent if $\varepsilon = 0$. We say $\mathcal{G} : \{0,1\}^\ell \to [m_0] \times \cdots \times [m_{d-1}]$ is an (ε-)pairwise independent generator if its outputs are (ε-)pairwise independent given a uniformly distributed seed.

Lemma 2.2. *Suppose $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{d-1})$ is a pairwise independent generator where $\mathcal{G}_i : \{0,1\}^\ell \to [m]$. Define $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{d-1})$ where $\mathcal{G}'_i(x) = \mathcal{G}_i(x) \bmod m_i$ for $i \in [d]$. Then $\mathcal{G}' : \{0,1\}^\ell \to [m_0] \times \cdots \times [m_{d-1}]$ is an ε-pairwise independent generator where $\varepsilon = 2/m$.*

Proof. For distinct $i, j \in [d]$ and $x \in [m_i]$, $x' \in [m_j]$, let B (resp. B') be the preimages of x (resp. x') under the map $s \mapsto s \bmod m_i$ (resp. $s \mapsto s \bmod m_j$). Then $||B| - m/m_i| \leqslant 1$ and $||B'| - m/m_j| \leqslant 1$. So $\mathbf{Pr}_s[\mathcal{G}'_i(s) = x] = |B|/m$ which differs from $1/m_i$ by at most $1/m$. Similarly $\mathbf{Pr}_s[\mathcal{G}'_i(s) = x \wedge \mathcal{G}'_j(s) = x'] = |B||B'|/m^2$ which differs from $1/(m_i m_j)$ by at most $2/m$. ∎

PRGs for Combinatorial Rectangles. Given $d \in \mathbb{N}$ and finite set $S = S_0 \times \cdots \times S_{d-1}$, let

$$\mathcal{M}_{S,d} \triangleq \{A_0 \times \cdots \times A_{d-1} : A_0 \subseteq S_0, \ldots, A_{d-1} \subseteq S_{d-1}\} = \prod_{i \in [d]} \mathcal{P}(S_i),$$

where $\mathcal{P}(S_i)$ is the power set of set S_i. The members of $\mathcal{M}_{S,d}$ are called (S,d)-combinatorial rectangles.

Definition 2.3 (PRGs for Combinatorial Rectangles)**.** *Given $\varepsilon > 0$, $d \in \mathbb{N}$ and finite set $S = S_0 \times \cdots \times S_{d-1}$, suppose a function $f : \{0,1\}^\ell \to S_0 \times \cdots \times S_{d-1}$ satisfies the following property*

$$\left|\mathbf{Pr}_{x \in \{0,1\}^\ell}\left[\, f(x) \in A \,\right] - \mathbf{Pr}_{x \in S_0 \times \cdots \times S_{d-1}}\left[\, x \in A \,\right]\right| \leqslant \varepsilon$$

for all $A \in \mathcal{M}_{S,d}$, then we say that f is a PRG that ε-fools $\mathcal{M}_{S,d}$ with seed length ℓ.

Let $S = [m]^d$. By probabilistic methods there exists a PRG ε-fooling $\mathcal{M}_{S,d}$ with seed length $O(\log m + \log d + \log(1/\varepsilon))$. The problem of explicitly constructing PRGs matching this bound is currently open. There is a long line of research devoting to this problem [3, 17, 25, 30]. The current state of the art is given by [25], which explicitly constructs PRGs ε-fooling $\mathcal{M}_{S,d}$ with seed length $O(\log m + \log d) + \widetilde{O}(\log(1/\varepsilon))$.

Theorem 2.4 ([25])**.** *Let $S = [m]^d$. There exists an explicit PRG $\mathcal{G}$ that ε-fools $\mathcal{M}_{S,d}$ with seed length $O(\log m + \log d + \log(1/\varepsilon) \log\log(1/\varepsilon) \log\log\log(1/\varepsilon))$.* [1]

[1] In [25] the seed length is presented as $O((\log\log m)(\log m + \log d + \log(1/\varepsilon))) + \widetilde{O}(\log(1/\varepsilon))$. But there are techniques of reducing m and d to $m' = (1/\varepsilon)^{O(1)}, d' = (1/\varepsilon)^{O(1)}$ using $O(\log m + \log d)$ randomness, cf. [3, 30].

Note that the seed length of the above PRG is $O(\log m)$ if d is polynomial in m and $1/\varepsilon$ is slightly sub-polynomial in m.

We then prove an analogue of Lemma 2.2.

Lemma 2.5. *Suppose $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{d-1})$ is a PRG ε-fooling $\mathcal{M}_{S,d}$ where $S = [m]^d$. Define $\mathcal{G}' = (\mathcal{G}'_0, \ldots, \mathcal{G}'_{d-1})$ where $\mathcal{G}'_i(x) = \mathcal{G}_i(x) \bmod m_i$ for $i \in [d]$. Then $\mathcal{G}'$ is a PRG ε'-fooling $\mathcal{M}_{S',d}$, where $S' = [m_0] \times \cdots \times [m_{d-1}]$, and $\varepsilon' = \varepsilon + (m_0 + \cdots + m_{d-1})/m$.*

Proof. For $A_0 \times \cdots \times A_{d-1} \in \mathcal{M}_{S',d}$, let B_i be the preimages of A_i under the map $s \mapsto s \bmod m_i$ for all i. Then $B_0 \times \cdots \times B_{d-1} \in \mathcal{M}_{S,d}$. Then $\mathbf{Pr}_s\left[\bigwedge_{i\in[d]} \mathcal{G}'_i(s) \in A_i\right] = \mathbf{Pr}_s\left[\bigwedge_{i\in[d]} \mathcal{G}_i(s) \in B_i\right]$ which differs from $\prod_{i\in[d]}(|B_i|/m)$ by at most ε since $\mathcal{G}$ is a pseudorandom generator ε-fooling $\mathcal{M}_{S,d}$. Note that $||B_i| - |A_i| \cdot m/m_i| \leqslant |A_i| \leqslant m_i$ for all i. A simple induction shows that $\prod_{i\in[d]}(|B_i|/m)$ differs from $\prod_{i\in[d]}(|A_i|/m_i)$ by at most

$$\sum_{i\in[d]} ||B_i|/m - |A_i|/m_i| \leqslant (m_0 + \cdots + m_{d-1})/m. \qquad \blacksquare$$

PRGs for Branching Programs

Definition 2.6 (Branching Programs)**.** *A branching program of length L, width W and degree D, or an (L, W, D)-branching program, is a directed (multi)-graph with node set $[W]\times\{0, \ldots, L\}$. We say the nodes in $[W]\times\{i\}$ are on the ith layer for $0 \leqslant i \leqslant L$. Each node (u, i) except those on the last layer has D outgoing edges to nodes on the next layer, and these D edges are associated with D distinct labels from $[D]$.*

Let $\mathcal{B}$ be a branching program of length L, width W and degree D. For $x = (x_1, \ldots, x_L) \in [D]^L$ and a node $(s, 0)$ on the first layer, define $\mathcal{B}(s, x) \in [n]$ such that the random walk that starts from $(s, 0)$ and takes the edge with label x_i at the ith step for $1 \leqslant i \leqslant L$ finally arrives at $(\mathcal{B}(s, x), L)$.

Definition 2.7 (PRGs for branching programs)**.** *A function $f : \{0,1\}^\ell \to [D]^L$ is a PRG that ε-fools (L, W, D)-branching programs if for any branching program $\mathcal{B}$ of length L, width W and degree D, any node $(s, 0)$ on the first layer, and any (t, L) on the last layer, it holds that*

$$\left|\mathbf{Pr}_{x\in\{0,1\}^\ell}[\,\mathcal{B}(s, f(x)) = t\,] - \mathbf{Pr}_{x\in[D]^L}[\,\mathcal{B}(s, x) = t\,]\right| \leqslant \varepsilon.$$

Many attempts were made to construct explicit PRGs for branching programs [29, 32, 33], due to its connection with the **RL** vs. **L** problem. The probabilistic method guarantees the existence of a (non-explicit) PRG that ε-fools (L, W, D)-branching programs with seed length $O(\log L + \log W + \log D + \log(1/\varepsilon))$. The problem of explicitly constructing such PRGs remains a great open problem. See [29, 32, 33] for known explicit constructions. We will use the constructions in [29, 32].

Theorem 2.8 ([29, 32])**.** *There exists an explicit PRG f that ε-fools (L, W, D)-branching programs with seed length $O(\log L(\log W + \log D + \log L + \log(1/\varepsilon)))$.*

Combining different generators. Finally, we show the properties of the combination of different PRGs.

Lemma 2.9. *There exists an explicit function $\mathcal{G} : \{0,1\}^\ell \to [m]^d$ that is both a pairwise independent generator and a PRG ε-fooling $\mathcal{M}_{[m]^d,d}$ with seed length*

$$O(\log m + \log d + \log(1/\varepsilon) \log\log(1/\varepsilon) \log\log\log(1/\varepsilon)).$$

Indeed, suppose $\mathcal{G}^\flat = (\mathcal{G}^\flat_0, \ldots, \mathcal{G}^\flat_{d-1})$ *is a pairwise independent generator with seed length* $\ell_1 = O(\log m + \log d)$*, where* $\mathcal{G}^\flat_i : \{0,1\}^{\ell_1} \to [m]$ *for* $i \in [d]$*. Suppose* $\mathcal{G}^\sharp = (\mathcal{G}^\sharp_0, \ldots, \mathcal{G}^\sharp_{d-1})$ *is a* PRG *ε-fooling* $\mathcal{M}_{[m]^d,d}$ *with seed length* $\ell_2 = O(\log m + \log d + \log(1/\varepsilon) \log\log(1/\varepsilon) \log\log\log(1/\varepsilon))$*. Identify* $[m]$ *with* $\mathbb{Z}_m$ *and define* $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{d-1}) : \{0,1\}^{\ell_1} \times \{0,1\}^{\ell_2} \to [m]^d$ *by*

$$\mathcal{G}_i(x_1, x_2) = \mathcal{G}^\flat_i(x_1) + \mathcal{G}^\sharp_i(x_2)$$

for $i \in [d]$*, where the additions are performed over* $\mathbb{Z}_m$*. Then* $\mathcal{G}$ *is both a pairwise independent generator and a* PRG *ε-fooling* $\mathcal{M}_{[m]^d,d}$*.*

Proof. Fix $x_1 \in \{0,1\}^{\ell_1}$ and consider x_2 uniformly distributed over $\{0,1\}^{\ell_2}$. Let $A_0 \times \cdots \times A_{d-1} \in \mathcal{M}_{[m]^d,d}$ be a combinatorial rectangle. For $i \in [d]$, let $A'_i = \{y - \mathcal{G}^\flat_i(x_1) : y \in A_i\}$ be the set obtained by shifting A_i by $(-\mathcal{G}^\flat_i(x_1))$ in $[m] = \mathbb{Z}_m$. Hence the probability that $\mathcal{G}(x_1, x_2)$ is in a combinatorial rectangle $A_0 \times \cdots \times A_{d-1} \in \mathcal{M}_{[m]^d,d}$ is just the probability that $\mathcal{G}^\sharp(x_2)$ is in $(A'_0 \times \cdots \times A'_{d-1}) \in \mathcal{M}_{[m]^d,d}$. This probability differs from $\prod_{i\in[d]}(|A'_i|/m) = \prod_{i\in[d]}(|A_i|/m)$ by at most ε since $\mathcal{G}^\sharp$ ε-fools $\mathcal{M}_{[m]^d,d}$. So $\mathcal{G}$ ε-fools $\mathcal{M}_{[m]^d,d}$ as well. A similar argument shows that $\mathcal{G}$ is a pairwise independent generator. ∎

2.2 Unbalanced Expanders with Near-Optimal Expansion

We consider the following kind of left-regular bipartite graphs.

Definition 2.10. *Let* $\Gamma : [N] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$ *be a function where* $\Gamma(x,y) \in [M_y]$ *for any* $x \in [N]$*,* $y \in [D]$*. Function* Γ *specifies a left-degree* D *bipartite graph with left vertex set* $[N]$ *and right vertex set* $\bigsqcup_{i\in[D]}[M_i]$ *in the following way: for* $x \in [N]$ *and* $y \in [D]$*, the* y*th neighbor of* x *is given by* $\Gamma(x,y)$*.*

We are interested in graphs Γ exhibiting excellent expansion properties. This leads to the notion of unbalanced expanders [26, 34].

Definition 2.11 (Unbalanced expanders [26, 34])**.** *Let* $\Gamma : [N] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$ *be as in Definition 2.10. We call* Γ *a* (K, A)*-expander if for any set* $S \subseteq [N]$ *of size* K*, it holds that* $|N(S)| \geqslant AK$*. We call* Γ *a* $(\leqslant K, A)$*-expander if it is a* (K', A)*-expander for all* $K' \leqslant K$*.* [2]

In particular we are interested in (K, A)-expanders, where the parameter $A = (1-\varepsilon)D$ for small ε, i.e. for any subset S of size K from the left set $[N]$, there is almost no collision among the neighbors of nodes in S. Explicit constructions of such unbalanced expanders with near-optimal expansion are known.

Theorem 2.12 ([26])**.** *For any* $N \in \mathbb{N}$*,* $K \leqslant N$*, and* $\varepsilon > 0$*, there is an explicit* $(K, (1-\varepsilon)D)$*-expander* $\Gamma : [N] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$ *with* $D = \left(\frac{\log N}{\varepsilon}\right)^{O(1)}$ *and* $M_0 = \cdots = M_{D-1} \leqslant \max\left\{D, K^{O(1)}\right\}$*.*

Assume that $\Gamma : [N] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$ is a $(K, (1-\varepsilon)D)$-expander. We consider the map $\Gamma(\cdot, U)$ applied on any K elements of $[N]$ where U is uniformly distributed over $[D]$. The following lemma states that with high probability these K elements are mapped into $\bigsqcup_{i\in[D]}[M_i]$ with almost no collision.

Lemma 2.13. *Let* $\Gamma : [N] \times [D] \to \bigsqcup_{i\in[D]}[M_i]$ *be a* $(K, (1-\varepsilon)D)$*-expander. Let* S *be a subset of* $[N]$ *of size* K*. Then for at least* $(1-\sqrt{\varepsilon})$*-fraction of* $y \in [D]$*, it holds that* $|\{\Gamma(x,y) : x \in S\}| \geqslant (1-\sqrt{\varepsilon})K$*.*

[2] The definition here is slightly different from [26, 34] as we require $\Gamma(x,y) \in [M_y]$. This is analogous to the difference between standard and strong condensers.

Proof. The size of $N(S) = \bigsqcup_{y\in[D]}\{\Gamma(x,y) : x \in S\}$ is at least $(1-\varepsilon)DK$ as Γ is a $(K,(1-\varepsilon)D)$-expander. So $\mathbf{E}_y[|\{\Gamma(x,y) : x \in S\}|] \geqslant (1-\varepsilon)K$ with y uniformly distributed over $[D]$. Also note that $|\{\Gamma(x,y) : x \in S\}| \leqslant |S| = K$ for any $y \in [D]$. Applying Markov's inequality on $K - |\{\Gamma(x,y) : x \in S\}|$, we have $\mathbf{Pr}_y[|\{\Gamma(x,y) : x \in S\}| < (1-\sqrt{\varepsilon})K] \leqslant \sqrt{\varepsilon}$. ■

3 Protocol for Expander Graphs

Let $G = (V, E)$ be an undirected and simple graph with $V[G] = [n]$. We consider only T'-round protocols for G, in which nodes send rumors only for the first T' rounds, and assume that $T' = (\log n)^{O(1)}$ is an upper bound of the rumor spreading time. We combine the protocols with an ID distribution mechanism so that every node gets a unique ID once it gets the rumor. Specially, in round 0 there is one arbitrary node having the rumor, and the ID of this node is set to be 0. We assume that node 0 knows T', n, and the maximum degree Δ. Moveover, node 0 chooses a binary string, called *seed*, uniformly at random, and the seed is appended to the rumor. In subsequent rounds, whenever one node with ID u sends the rumor to one of its neighbors in round t, it also sends a unique string consisting of the ID u, current round number t, as well as parameter T', n and Δ (Our protocols work fine as long as the initial node having the rumor knows some upper bound n' of n, where n' is a polynomial of n). A node is *uninformed* as long as it has not received a rumor. Once a node receives the first rumor from an informed node with ID u in round t, it becomes *informed* and gets a unique ID defined by $g_t(u) \triangleq 2^{t-1} + u$. If one node becomes informed from multiple informed nodes, then this node chooses an arbitrary node with ID u that informs it and uses $g_t(u)$ as its ID.

Claim 3.1 ([24])**.** *Through the protocol above, all informed nodes have different IDs. Moreover, if the protocol finishes in T rounds, then all the IDs are in $\left[2^T\right]$.*

Throughout the rest of the paper we assume that all the IDs are in a set $[0, n^c]$ for a constant $c \geqslant 1$. Denote by s the initial node having the rumor. All the omitted proofs of the lemmas can be found in the appendix.

3.1 Protocol

Let $\mathbf{A}_G$ be the adjacency matrix of G, and $\mathbf{D}$ the $n \times n$ diagonal matrix defined by $\mathbf{D}_{uu} = \deg(u)$ for $u \in V[G]$. Let $\mathbf{M}_G = \mathbf{D}^{-1}\mathbf{A}_G$ be the transition matrix for the random walk over G, and

$$\mathbf{N}_G \triangleq \mathbf{D}^{-1/2}\mathbf{A}_G\mathbf{D}^{-1/2} = \mathbf{D}^{1/2}\mathbf{M}_G\mathbf{D}^{-1/2}.$$

Define the n real eigenvalues of $\mathbf{N}_G$ by $1 = \lambda_1 \geqslant \ldots \geqslant \lambda_n \geqslant -1$.

Definition 3.2 (Spectral Expansion)**.** *Graph G has spectral expansion α if $\lambda \leqslant 1-\alpha$, where λ is the second largest (absolute) eigenvalue of $\mathbf{N}_G$.*

In this subsection we study protocols for graphs G with $\Delta/\delta = O(1)$ and spectral expansion α. Let $\beta \triangleq \Delta/\delta = O(1)$. The main result of Section 3 is as follows.

Theorem 3.3. *Let G be a graph of n nodes with $\Delta/\delta = O(1)$ and spectral expansion α satisfying $\alpha \geqslant 8\Delta^2/\delta^{2.5}$. Then the following statements hold: (1) There is a protocol using $O(\log n(\log\log n + \log(1/\alpha)))$ random bits such that with high probability, all but $o(1)$-fraction of the nodes in G get informed in $O(\alpha^{-2}\log n)$ rounds. (2) Moreover, if $\delta = \Theta(n)$, then there is an explicit protocol using $O(\log n(\log\log n + \log(1/\alpha)))$ random bits such that with high probability all nodes in G get informed in $O(\alpha^{-2}\log n)$ rounds.*

For the case that α is a positive constant, i.e. G is a expander graph, we may always assume that $\alpha \geqslant 8\Delta^2\delta^{-2.5}$, since otherwise $\Delta < 64\beta^5/\alpha^2 = O(1)$ and there is a simple deterministic protocol that finishes in $O(\log n)$ rounds.

Corollary 3.4. *Let G be an expander graph of n nodes where $\Delta/\delta = O(1)$, $\delta = \omega(1)$ and $\alpha = \Theta(1)$. Then the following statements hold: (1) There is a protocol using $O(\log n \log\log n)$ random bits such that with high probability, all but $o(1)$-fraction of the nodes in G get informed in $O(\log n)$ rounds. (2) If G is a complete graph of n nodes, then there is an explicit protocol using $O(\log n \log\log n)$ random bits such that with high probability all nodes in G get informed in $O(\log n)$ rounds.*

This result is based on a reduction between rumor spreading protocols and PRGs for branching programs. We remark that the randomness complexity for Theorem 3.3 and Corollary 3.4 is sub-optimal only because we do not yet know good enough explicit PRGs for branching programs and, as stated later, any construction of the PRGs with optimal parameters implies the explicit construction of randomness-optimal rumor spreading protocols for dense graphs of n nodes that use $\Theta(\log n)$ random bits in total. See Theorem C.2 for the lower bound of randomness requirement for graph G with $\delta = \Theta(n)$.

Our protocol is based on pairwise independent generators, and PRGs for branching programs, and the formal description of the protocol is as follows:

Protocol 1 (Protocol for Expander Graphs)**.** *Let $m = n^{\Theta(1)}$ be a sufficiently large power of 2. Pick the following objects:*

- *An explicit pairwise independent generator $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}) : \{0,1\}^{\ell} \to [m]^n$, where $\ell = O(\log n)$.*
- *An explicit PRG $f = (f_1, \ldots, f_T) : \{0,1\}^{\ell'} \to \left(\{0,1\}^{\ell}\right)^T$ that ε-fools $(T, n^2, 2^{\ell})$-branching programs with $\varepsilon = n^{-3}$ and seed length $\ell' = O(\log n \log T)$.*

These two objects $\mathcal{G}$ and f can be uniquely constructed from n^c, and hence are known to every informed node.

The initial node having the rumor chooses a random string $x \in \{0,1\}^{\ell'}$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, $1 \leqslant i \leqslant T$, node u sends the rumor to the neighbor with index $\mathcal{G}_u(f_i(x)) \bmod \deg(u)$ in its adjacency list.

Proposition 3.5. *Assume that Protocol 1 finishes in T rounds. Then it uses $O(\log n \log T)$ random bits in total.*

Remark 3.6. *One difference between Protocol 1 and the protocols in [24] is that in Protocol 1 different rounds use one common seed x, while in [24] the seeds used by different rounds are different and mutually independent, and hence the analysis is much easier.*

3.2 Analysis

We introduce the notation used in the analysis. For $m \in \mathbb{N}$, vector $\mathbf{u} \in \mathbb{R}^m$ and real number $p \geqslant 1$, define the norm $\|\mathbf{u}\|_p = (\sum_{i=1}^m |\mathbf{u}_i|^p)^{1/p}$. In particular, $\|\mathbf{u}\|_\infty = \max_{1 \leqslant i \leqslant m} |\mathbf{u}_i|$. The inner product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ is $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^m \mathbf{u}_i \mathbf{v}_i$. We write $\mathbf{1}_m$ for the vector in $\mathbb{R}^m$ having ones in all entries, or simply $\mathbf{1}$ if the dimension is clear from the context. Similarly write $\mathbf{0}_m$ or $\mathbf{0}$ for the zero vector. Let $\mathbf{e}_i$ be the vector that has an one in the ith entry and zero elsewhere. Write $\mathbf{I}_m$ or $\mathbf{I}$ for the $m \times m$ identity matrix. For a matrix $\mathbf{M} \in \mathbb{R}^{m \times m'}$, we use $\mathbf{M}_{ij}$ to denote the entry on $\mathbf{M}$'s ith row and jth column. For $p \in [1, \infty) \cup \{\infty\}$, define

$$\|\mathbf{M}\|_p = \sup_{\mathbf{u} \in \mathbb{R}^m \setminus \{\mathbf{0}\}} \frac{\|\mathbf{u}\mathbf{M}\|_p}{\|\mathbf{u}\|_p}.$$

It is easy to show that $\|\mathbf{M}\|_1$ equals the maximum of the ℓ_1-norms of the rows of $\mathbf{M}$. And $\|\mathbf{M}\|_\infty$ equals the maximum of the ℓ_1-norms of the columns of $\mathbf{M}$, or equivalently $\|\mathbf{M}^\intercal\|_1$. We say a square matrix $\mathbf{M}$ is stochastic if all of its entries are non-negative and all of its rows have ℓ_1-norm 1. Clearly if $\mathbf{M}$ is stochastic, then $\|\mathbf{M}\|_1 = 1$.

Lemma 3.7. $\|\mathbf{M}\|_2^2 \leqslant \|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty$.

Define linear operators $\mathcal{T}_1, \mathcal{T}_2 : \mathbb{R}^{n^2} \to \mathbb{R}^n$ such that, for a vector $\mathbf{u} \in \mathbb{R}^{n^2}$ indexed by $(u, v) \in V[G] \times V[G]$, we have $(\mathcal{T}_1(\mathbf{u}))_u = \sum_{v \in V[G]} \mathbf{u}_{(u,v)}$ and $(\mathcal{T}_2(\mathbf{u}))_u = \sum_{v \in V[G]} \mathbf{u}_{(v,u)}$. When $\mathbf{u}$ represents a probability distribution over $V[G] \times V[G]$, the vectors $\mathcal{T}_1(\mathbf{u})$ and $\mathcal{T}_2(\mathbf{u})$ represents its two marginal distributions.

3.2.1 Proof Sketch

Now we discuss the intuitions behind constructing the protocol and sketch our proof. A main technique that we use here is a generic reduction between rumor spreading processes and branching programs. More specifically, we compare the process of rumor spreading with a random walk on a branching program. For random walks, a walk always stays at a single node throughout the process, although this node keeps changing. On the other hand, in the process of rumor spreading, each informed node u randomly sends the rumor to one of its neighbors v in each round, and then u, v are both informed subsequently. So we may think of rumor spreading as many random walks in parallel: When node u sends the rumor to v, one random walk moves from u to v whereas another one stays at u. In order to characterize this behavior precisely, we introduce the notion of colored random walks.

Definition 3.8 (Colored Random Walk)**.** *Given a random rumor spreading process on a graph G with the initial node s, an associated colored random walk of length T is a sequence of $T+1$ nodes $(p_0, \ldots, p_T)$ of G together with a lazy/non-lazy coloring on the pairs (p_i, p_{i+1}) called colored edges for $0 \leqslant i < T$, such that $p_0 = s$, and*

1. *if (p_i, p_{i+1}) is lazy, then $p_i = p_{i+1}$;*
2. *if (p_i, p_{i+1}) is non-lazy, then p_i picks p_{i+1} in round $i+1$ in the rumor spreading process.*

Node s itself is a colored random walk of length 0.

We denote by $\mathcal{C}_T \triangleq \{\text{lazy}, \text{non-lazy}\}^T$ the set of possible color patterns for length-T colored paths. For $S \in \mathcal{C}_T$, there exists a unique colored random walk of length T with color pattern S associated with a given instantiation of a rumor spreading process. Let X_u^S be the indicator random variable of the event that the colored random walk with color pattern S finally reaches node u. Note that u receives the rumor in T rounds if some colored random walk of length T arrives at u. Therefore

$$\mathbf{Pr}\left[\, u \text{ receives the rumor in } T \text{ rounds} \,\right] = \mathbf{Pr}\left[\sum_{S \in \mathcal{C}_T} X_u^S > 0 \right]. \tag{3.1}$$

We want to reduce the global event $\sum_{S \in \mathcal{C}_T} X_u^S > 0$ to local events $X_u^S > 0$. Note that the indicator of $\sum_{S \in \mathcal{C}_T} X_u^S > 0$ is the conjunction of those of $X_u^S > 0$. We use the Cauchy-Schwarz inequality to "linearize" the conjunction operation and then use the linearity of expectations:

$$\begin{aligned}
\mathbf{Pr}\left[\sum_{S \in \mathcal{C}_T} X_u^S > 0 \right] = \mathbf{E}\left[\mathbf{1}_{\sum_{S \in \mathcal{C}_T} X_u^S > 0} \right] &\geqslant \frac{\mathbf{E}\left[\sum_{S \in \mathcal{C}_T} X_u^S \right]^2}{\mathbf{E}\left[\left(\sum_{S \in \mathcal{C}_T} X_u^S \right)^2 \right]} \\
&= \frac{\sum_{S, S' \in \mathcal{C}_T} \mathbf{E}\left[X_u^S \right] \mathbf{E}\left[X_u^{S'} \right]}{\sum_{S, S' \in \mathcal{C}_T} \mathbf{E}\left[X_u^S X_u^{S'} \right]}.
\end{aligned} \tag{3.2}$$

Note that each $\mathbf{E}\left[X_u^S \right]$ only depends on a single random walk with color pattern S, and hence can be well approximated by the PRGs for branching programs. The same is true for each

$\mathbf{E}\left[X_u^S X_u^{S'} \right]$ if we use branching programs with set of states $V[G] \times V[G]$ to keep track of two colored random walks simultaneously.

If colored random walks were pairwise independent, then $\mathbf{E}\left[X_u^S X_u^{S'} \right] = \mathbf{E}\left[X_u^S \right] \mathbf{E}\left[X_u^{S'} \right]$ for $S \neq S'$ and the lower bound given by (3.2) would be very close to 1. Unfortunately they are not pairwise independent: if two colored paths are at the same node in some round and their next steps are both non-lazy, they will move to the same random neighbor rather than two independent random neighbors. Consequently, for a typical pair of colored random walks (S, S') we only have $\mathbf{E}\left[X_u^S X_u^{S'} \right] \leqslant c \cdot \mathbf{E}\left[X_u^S \right] \mathbf{E}\left[X_u^{S'} \right]$ where $c > 1$ is some constant. Then the probability lower bound given by (3.2) is no better than $1/c$. To remedy this problem, we observe that (3.2) still holds if we replace $\mathcal{C}_T$ by any subset $\mathcal{C}'_T \subseteq \mathcal{C}_T$, and (3.1) can be rewritten as

$$\mathbf{Pr}\left[u \text{ receives the rumor in } T \text{ rounds} \right] \geqslant \mathbf{Pr}\left[\sum_{S \in \mathcal{C}'_T} X_u^S > 0 \right].$$

for any $\mathcal{C}'_T \subseteq \mathcal{C}_T$. This leads to the following lemma.

Lemma 3.9. *In the random rumor spreading process, the probability that node $u \in V[G]$ receives the rumor in T rounds is lower bounded by*

$$\frac{\sum_{S,S' \in \mathcal{C}'_T} \mathbf{E}\left[X_u^S \right] \mathbf{E}\left[X_u^{S'} \right]}{\sum_{S,S' \in \mathcal{C}'_T} \mathbf{E}\left[X_u^S X_u^{S'} \right]},$$

where $\mathcal{C}'_T$ is an arbitrary subset of $\mathcal{C}_T$.

By a careful choice of $\mathcal{C}'_T$ along with a delicate analysis we can show that two typical colored random paths are "mildly" pairwise independent, yielding a $1 - o(1)$ lower bound in Lemma 3.9. Hence the construction of PRGs for branching programs implies an explicit protocol for rumor spreading.

Another important issue is that random walks on branching programs only take one step each time, whereas in the rumor spreading process multiple steps are took simultaneously, each requiring some randomness. We use a pairwise independent generator for the choices of these steps. This is perfectly compatible with the discussion above as we only use expectations of monomials X_u^S and $X_u^S X_u^{S'}$, neither involving more than two variables and hence pairwise independence suffices. Formally the following theorem shows a reduction from PRGs for branching programs to rumor spreading protocols.

Theorem 3.10. *Suppose f is an explicit PRG that ε-fools (L, W, D)-branching programs with seed length ℓ, where $L = \Theta(\alpha^{-2} \log n)$, $D = n^{\Theta(1)}$ sufficiently large, $\varepsilon = n^{-3}$, and $W = n^2$. Then there exists an explicit rumor spreading protocol using ℓ random bits such that, for any graph G with n nodes, $\Delta/\delta = O(1)$ and spectral expansion $\alpha \geqslant 8\Delta^2\delta^{-2.5}$, all but $o(1)$ fraction of the nodes in G get informed in $O(\alpha^{-2} \log n)$ rounds with high probability.*

By Theorem 3.10, an explicit PRG construction with good enough parameters, in particular with seed length $\ell = O(\log n)$ and error $\varepsilon = n^{-3}$, yields explicit rumor spreading protocols using only $O(\log n)$ random bits. In addition, we usually only need to fool short branching programs instead of those in full generality (i.e. of length $n^{\Theta(1)}$) since the process of rumor spreading typically finishes in a small number of rounds. The PRG in [33] fools such branching programs using only $O(\log n)$ random bits, but the error obtained is too large for our application, being $2^{-\log^{1-\nu} n}$ for any constant $\nu > 0$ instead of $n^{-\Theta(1)}$. We will use the PRG constructions from [29] or [32] instead for our protocol.

Theorem 3.11 ([29, 32])**.** *There exists an explicit PRG f that ε-fools (L, W, D)-branching programs with seed length $O(\log n \log L)$ for $W = n^{\Theta(1)}, D = n^{\Theta(1)}$, and $1/\varepsilon = n^{\Theta(1)}$.*

The first statement of Theorem 3.3 follow from Theorem 3.10 and Theorem 3.11.

3.2.2 Proof of Theorem 3.10

To prove Theorem 3.10, we first generalize Protocol 1 to a family of (not necessarily explicit) protocols. That is, instead of studying Protocol 1 where every informed node chooses neighbors according to the output of the PRGs, we consider a family of protocols, and each of them is specified by a joint distribution of random variables $\mathcal{P} = (X_1, \ldots, X_T)$, such that the random choices in round i are determined by the ith variable X_i with a pairwise independent generator.

Protocol 2 (Protocol via distribution $\mathcal{P}$)**.** *Let $m = n^{\Theta(1)}$ be a sufficiently large power of 2. Let $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}) : \{0,1\}^\ell \to [m]^n$ be an explicit pairwise independent generator, where $\ell = O(\log n)$. A distribution $\mathcal{P}$ over $\left(\{0,1\}^\ell\right)^T$ specifies the following protocol:*

The initial node having the rumor chooses a random string $(x_1, \ldots, x_T) \in \left(\{0,1\}^\ell\right)^T$ according to the distribution $\mathcal{P}$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, node u chooses the neighbor with index $\mathcal{G}_u(x_i) \bmod \deg(u)$ in its adjacency list to send the rumor.

Setting $\mathcal{P}$ as the uniform distribution, denoted by $\mathcal{U}$, yields the hashing-based protocol in [24]. We recover Protocol 1 by letting $\mathcal{P}$ be the output of a PRG for branching programs.

We wish to bound $\mathbf{E}\left[X_u^S\right]$ and $\mathbf{E}\left[X_u^S X_u^{S'}\right]$ and then apply Lemma 3.9. For the family of protocols above, these quantities can be easily characterized using branching programs. For color patterns $S, S' \in \mathcal{C}_T$, define the branching program $\mathcal{B}_{S,S'}$ of length T, width n^2 and degree 2^ℓ as follows: For $(u,v) \in V[G] \times V[G]$, $i \in [T]$ and $x \in \{0,1\}^\ell$, the edge of $((u,v),i)$ labeled x goes to $((w,y), i+1)$ where w and y are picked in the following way: If the ith color of S (resp. S') is lazy, then let $w = u$ (resp. $y = v$), otherwise let w (resp. y) be the zth neighbor of u (resp. v), where $z = \mathcal{G}_u(x) \bmod \deg(u)$ (resp. $\mathcal{G}_v(x) \bmod \deg(v)$). The following lemma follows immediately from this construction:

Lemma 3.12. *For Protocol 2 that is specified by a distribution $\mathcal{P}$ over $\left(\{0,1\}^\ell\right)^T$ and any $u, v \in V[G]$, $S, S' \in \mathcal{C}_T$, it holds that*

$$\mathbf{E}\left[X_u^S X_v^{S'}\right] = \mathbf{Pr}_{x\sim\mathcal{P}}\left[\mathcal{B}_{S,S'}((s,s),x) = (u,v)\right]$$

and

$$\mathbf{E}\left[X_u^S\right] = \sum_{v\in V[G]} \mathbf{Pr}_{x\sim\mathcal{P}}\left[\mathcal{B}_{S,S'}((s,s),x) = (u,v)\right] = \sum_{v\in V[G]} \mathbf{Pr}_{x\sim\mathcal{P}}\left[\mathcal{B}_{S',S}((s,s),x) = (v,u)\right].$$

Recall that $\mathbf{M}_G = \mathbf{D}^{-1}\mathbf{A}_G$ is the transition matrix for the random walk over G. We define by $\mathbf{P} \in \mathbb{R}^{n^2\times n^2}$ is the transition matrix with its rows and columns indexed by the set $V[G] \times V[G]$, which corresponds to two non-lazy steps in parallel:

$$\mathbf{P}_{(u,v)(w,x)} = \begin{cases} (\mathbf{M}_G \otimes \mathbf{M}_G)_{(u,v)(w,x)} & u \neq v, \\ (\mathbf{M}_G)_{(u,w)} & u = v, w = x, \\ 0 & u = v, w \neq x. \end{cases}$$

Remark 3.13. *Matrix $\mathbf{P}$ agrees with $\mathbf{M}_G \otimes \mathbf{M}_G$ except on the rows indexed by (u,u), $u \in V[G]$. This is a manifestation of the fact that the "non-lazy" steps from the same node made by two different colored random walks are not independent, i.e., every informed node can only send the rumor to one neighbor in each round.*

To study random walks associated with certain color patterns $S = (S_1, \ldots, S_T) \in \mathcal{C}_T, S' = (S'_1, \ldots, S'_T) \in \mathcal{C}_T$, we introduce the matrix $J^\star_{S,S'} \in \mathbb{R}^{n^2\times n^2}$ defined by

$$\mathbf{J}^\star_{S,S'} \triangleq \prod_{i=1}^{T} \mathbf{J}^{(i)}_{S,S'},$$

where each matrix $\mathbf{J}^{(i)}_{S,S'}$ is defined by

$$\mathbf{J}^{(i)}_{S,S'} = \begin{cases} \mathbf{I} \otimes \mathbf{I} & S_i = \text{lazy}, S'_i = \text{lazy} \\ \mathbf{I} \otimes \mathbf{M}_G & S_i = \text{lazy}, S'_i = \text{non-lazy} \\ \mathbf{M}_G \otimes \mathbf{I} & S_i = \text{non-lazy}, S'_i = \text{lazy} \\ \mathbf{P} & S_i = \text{non-lazy}, S'_i = \text{non-lazy}. \end{cases}$$

By definition, the stochastic matrix $\mathbf{J}^{\star}_{S,S'}$ corresponds to two colored random walks in parallel over the graph G, with the color patterns S and S' respectively.

The following lemma shows that the computation over the branching program $\mathcal{B}_{S,S'}$ can be characterized by the matrix $\mathbf{J}^{\star}_{S,S'}$, up to a negligible error term.

Lemma 3.14. *Let $\mathcal{U}$ be the uniform distribution over $\left(\{0,1\}^{\ell}\right)^T$. Then for any $u, v \in V[G]$, $S, S' \in \mathcal{C}_T$, it holds that*

$$\left|\mathbf{Pr}_{x\sim\mathcal{U}}\left[\,\mathcal{B}_{S,S'}((s,s),x) = (u,v)\,\right] - \left\langle \mathbf{e}_{(s,s)}\mathbf{J}^{\star}_{S,S'}, \mathbf{e}_{(u,v)}\right\rangle\right| \leqslant 2Tn^2/m.$$

Proof. Let $\mathbf{B}^{(i)} \in \mathbb{R}^{n^2\times n^2}$ be the transition matrix of the branching program $\mathcal{B}_{S,S'}$ from the $(i-1)$th to ith level. Then we have

$$\mathbf{Pr}_{x\sim\mathcal{U}}\left[\,\mathcal{B}_{S,S'}((s,s),x) = (u,v)\,\right] = \left\langle \mathbf{e}_{(s,s)}\prod_{i=1}^{T}\mathbf{B}^{(i)}, \mathbf{e}_{(u,v)}\right\rangle.$$

So it suffices to show that

$$\left\|\prod_{i=1}^{T}\mathbf{B}^{(i)} - \prod_{i=1}^{T}\mathbf{J}^{(i)}_{S,S'}\right\|_1 \leqslant 2Tn^2/m.$$

By the construction of $\mathcal{B}_{S,S'}$, a random walk at $((u_1,v_1),i-1)$ goes to $((u_2,v_2),i)$ where u_2 (resp. v_2) are picked as $\mathcal{G}_{u_1}(x) \bmod \deg(u_1)$ (resp. $\mathcal{G}_{v_1}(x) \bmod \deg(v_1)$) for random x (or $u_2 = u_1$ (resp. $v_2 = v_1$) if the corresponding step is lazy). Lemma 2.2 then implies that each entry $\mathbf{B}^{(i)}_{(u_1,v_1)(u_2,v_2)}$ differs from $\left(\mathbf{J}^{(i)}_{S,S'}\right)_{(u_1,v_1)(u_2,v_2)}$ by at most $2/m$, and hence

$$\left\|\mathbf{B}^{(i)} - \mathbf{J}^{(i)}_{S,S'}\right\|_1 \leqslant 2n^2/m$$

for all $1 \leqslant i \leqslant T$. A simple induction on T then shows

$$\left\|\prod_{i=1}^{T}\mathbf{B}^{(i)} - \prod_{i=1}^{T}\mathbf{J}^{(i)}_{S,S'}\right\|_1 \leqslant 2Tn^2/m.$$

■

Lemma 3.12 and Lemma 3.14 show that to bound $\mathbf{E}\left[\,X^S_u X^{S'}_v\,\right]$ and $\mathbf{E}\left[\,X^S_u\,\right]$, it suffices to study the stochastic matrices $\mathbf{J}^{\star}_{S,S'}$. Instead of analyzing $\mathbf{J}^{\star}_{S,S'}$ for different S, S' individually, we investigate them simultaneously for S, S' ranged over some set $\mathcal{C}'_T \subseteq \mathcal{C}_T$ defined as follows: For $k = 0, 1, \ldots, T/3$ (for simplicity assume T is a multiple of 3), define

$$\mathcal{C}'_{T,k} \triangleq \{(S_1,\ldots,S_T) \in \mathcal{C}_T : S_i = \text{lazy for } i \leqslant k \text{ or } i > 2T/3 + k\}$$

and let $\mathcal{C}'_T \triangleq \bigcup_{i=0}^{T/3}\mathcal{C}'_{T,i}$. Here $\mathcal{C}'_{T,k}$ consists of color patterns for colored random walks whose first k steps and last $T/3-k$ steps are lazy.

Let $\boldsymbol{\pi} \triangleq \mathbf{1D}/\|\mathbf{1D}\|_1$, and it is easy to see that $\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{M}_G$. Define vectors $\mathbf{x} \in \mathbb{R}^{n^2}$ and $\mathbf{y} \in \mathbb{R}^n$ such that

$$\mathbf{x}_{(u,v)} = \mathbf{E}_{S,S' \in \mathcal{C}'_T} \left[\left\langle \mathbf{e}_{(s,s)} \mathbf{J}^{\star}_{S,S'}, \mathbf{e}_{(u,v)} \right\rangle \right], \qquad \mathbf{y} = \mathcal{T}_1(\mathbf{x}). \tag{3.3}$$

Then the following key lemma states when the length of colored random walks is sufficiently large, $\mathbf{x}$ is close to the stationary distribution $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$ of $\mathbf{M}_G \otimes \mathbf{M}_G$, and similarly $\mathbf{y}$ is close to the stationary distribution $\boldsymbol{\pi}$ of $\mathbf{M}_G$. This implies that after sufficient long rounds, the probability that colored random walks hit arbitrary nodes can be approximated by the stationary distribution of the matrix $\mathbf{M}_G$, whose definition does not involve colored random walks.

Lemma 3.15. *$\|\mathbf{x} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 = o(n^{-3/2})$ and $\|\mathbf{y} - \boldsymbol{\pi}\|_2 = O(n^{-2})$ for sufficiently large $T = \Theta(\alpha^{-2} \log n)$.*

Lemma 3.15 is derived from a series of technical lemmas listed Section 3.2.3 below, whose proofs are deferred to the appendix. Now we are ready to prove Theorem 3.10.

Proof of Theorem 3.10. By Lemma 3.9 and Lemma 3.12 we have

$$\begin{aligned}
&\mathbf{Pr}\left[\,\text{u receives the rumor in } T \text{ rounds}\,\right] \\
&\geqslant \frac{\sum_{S,S' \in \mathcal{C}_T} \mathbf{E}\left[\, X_u^S \,\right] \mathbf{E}\left[\, X_u^{S'} \,\right]}{\sum_{S,S' \in \mathcal{C}_T} \mathbf{E}\left[\, X_u^S X_u^{S'} \,\right]} = \frac{\left(\sum_{v \in V[G]} \mathbf{Pr}_{x \sim \mathcal{P}}\left[\, \mathcal{B}_{S,S'}((s,s),x) = (u,v) \,\right]\right)^2}{\sum_{S,S' \in \mathcal{C}_T} \mathbf{Pr}_{x \sim \mathcal{P}}\left[\, \mathcal{B}_{S,S'}((s,s),x) = (u,v) \,\right]} = \frac{(\mathbf{y}_u)^2}{\mathbf{x}_{(u,u)}}.
\end{aligned}$$

Define $\mathbf{x}^{\mathcal{P}} \in \mathbb{R}^{n^2}$ and $\mathbf{y}^{\mathcal{P}} \in \mathbb{R}^n$ such that

$$\mathbf{x}^{\mathcal{P}}_{(u,v)} = \mathbf{E}_{S,S' \in \mathcal{C}'_T} \left[\mathbf{Pr}_{x \sim \mathcal{P}}\left[\, \mathcal{B}_{S,S'}((s,s),x) = (u,v) \,\right] \right], \qquad \mathbf{y}^{\mathcal{P}} = \mathcal{T}_1(\mathbf{x}'),$$

where $\mathcal{P}$ is the output of the PRG in Protocol 1. We define $\mathbf{x}'' \in \mathbb{R}^{n^2}$ and $\mathbf{y}'' \in \mathbb{R}^n$ in the same way except that $\mathcal{P}$ is replaced by $\mathcal{U}$.

As the distribution $\mathcal{P}$ is the output of the PRG, and hence ε-fools all branching programs $\mathcal{B}_{S,S'}$, we have $\left\|\mathbf{x}^{\mathcal{P}} - \mathbf{x}^{\mathcal{U}}\right\|_\infty \leqslant \varepsilon = n^{-3}$ and consequently $\left\|\mathbf{y}^{\mathcal{P}} - \mathbf{y}^{\mathcal{U}}\right\|_\infty \leqslant n\varepsilon = n^{-2}$.

By Lemma 3.14, we have $\left\|\mathbf{x}^{\mathcal{U}} - \mathbf{x}\right\|_\infty, \left\|\mathbf{y}^{\mathcal{U}} - \mathbf{y}\right\|_\infty = 2Tn^2/m = O(n^{-2})$ as $m = n^{\Theta(1)}$ is sufficiently large.

By Lemma 3.15, $\|\mathbf{y} - \boldsymbol{\pi}\|_\infty \leqslant \|\mathbf{y} - \boldsymbol{\pi}\|_2 = O(n^{-2})$. Therefore $\left\|\mathbf{y}^{\mathcal{P}} - \boldsymbol{\pi}\right\|_\infty = O(n^{-2})$. Note that $\boldsymbol{\pi}_u = \deg(u)/\sum_{u \in V[G]} \deg(u) \leqslant \beta n^{-1}$. Therefore $\mathbf{y}^{\mathcal{P}}_u/\boldsymbol{\pi}_u \geqslant 1 - O(n^{-1})$.

By Lemma 3.15, $\|\mathbf{x} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 = o(n^{-3/2})$. Define vector $\mathbf{w} \in \mathbb{R}^{n^2}$ such that $\mathbf{w}_{(u,v)} = 1$ if $u = v$ and $\mathbf{w}_{(u,v)} = 0$ otherwise. Then $\|\mathbf{w}\|_2 = n^{1/2}$. By the Cauchy-Schwarz inequality, we have

$$\mathbf{E}_{u \in V[G]}\left[\, \left|\mathbf{x}_{(u,u)} - (\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)}\right| \,\right] \leqslant n^{-1} \|\mathbf{x} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 \|\mathbf{w}\|_2 = o(n^{-2}).$$

By Markov's inequality, for all but $o(1)$-fraction of $u \in V[G]$, we have

$$\left|\mathbf{x}_{(u,u)} - (\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)}\right| = o(n^{-2})$$

and hence $\left|\mathbf{x}^{\mathcal{P}}_{(u,u)} - (\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)}\right| = o(n^{-2})$. For such u we have $\mathbf{x}^{\mathcal{P}}_{(u,u)}/(\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)} \leqslant 1 + o(1)$ since $(\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)} \geqslant \beta^{-2} n^{-2}$. Therefore, for all but $o(1)$-fraction of nodes $u \in V[G]$, the probability that u gets informed is at least

$$\frac{\left(\mathbf{y}^{\mathcal{P}}_u\right)^2}{\mathbf{x}^{\mathcal{P}}_{(u,u)}} \geqslant \frac{(1 - o(1))\,(\boldsymbol{\pi}_u)^2}{(1 + o(1))(\boldsymbol{\pi} \otimes \boldsymbol{\pi})_{(u,u)}} = 1 - o(1).$$

By Markov's inequality, with probability $1 - o(1)$, all but $o(1)$-fraction of the nodes get informed. ∎

3.2.3 Proof of Lemma 3.15

Let $1 = \lambda_1 \leqslant \ldots \leqslant \lambda_n \leqslant 1$ be the n eigenvalues of the matrix $\mathbf{N}_G = \mathbf{D}^{1/2}\mathbf{M}_G\mathbf{D}^{-1/2}$, associated with n normalized orthogonal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$. By the hypothesis of the spectral expansion of G, we know $\max\{|\lambda_2|, |\lambda_n|\} \leqslant 1 - \alpha$.

Note that $\|\mathbf{M}_G\|_1 = 1$ and $\|\mathbf{M}_G\|_\infty = \|\mathbf{D}^{-1}\mathbf{A}_G\|_\infty \leqslant \beta\|\Delta^{-1}\mathbf{A}_G\|_\infty \leqslant \beta$. By Lemma 3.7 we have $\|\mathbf{M}_G\|_2 \leqslant \sqrt{\beta}$. Similarly $\|\mathbf{M}_G \otimes \mathbf{M}_G\|_1 = 1$, $\|\mathbf{M}_G \otimes \mathbf{M}_G\|_\infty \leqslant \beta^2$, and $\|\mathbf{M}_G \otimes \mathbf{M}_G\|_2 \leqslant \beta$.

Define the matrix $\mathbf{Q} \in \mathbb{R}^{n^2 \times n^2}$ by

$$\mathbf{Q} \triangleq \frac{\mathbf{I} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{M}_G + \mathbf{M}_G \otimes \mathbf{I} + \mathbf{P}}{4}.$$

Note that matrix $\mathbf{Q}$ corresponds to two steps in parallel, each independently chosen as a lazy or non-lazy step with equal probability.

For a vector $\mathbf{u} \in \mathbb{R}^{n^2}$, we write $\mathbf{u} = \mathbf{u}^{\parallel} + \mathbf{u}^{\perp}$ as the (non-orthogonal) decomposition of $\mathbf{u}$ such that $\mathbf{u}^{\parallel} \parallel \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and $\mathbf{u}^{\perp} \perp \mathbf{1}$. The decomposition exists and is unique. Indeed, we have the following simple lemma:

Lemma 3.16. *For any vector $\mathbf{u} \in \mathbb{R}^{n^2}$, let $\mathbf{u} = \mathbf{u}^{\parallel} + \mathbf{u}^{\perp}$ be the (non-orthogonal) decomposition of $\mathbf{u}$ such that $\mathbf{u}^{\parallel} \parallel \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and $\mathbf{u}^{\perp} \perp \mathbf{1}$. Then the following statements hold:*

(a) Write $\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2} = \mathbf{u}_1 + \mathbf{u}_2$ such that $\mathbf{u}_1 \parallel \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$ and $\mathbf{u}_2 \perp \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$. Then $\mathbf{u}^{\parallel} = \mathbf{u}_1(\mathbf{D} \otimes \mathbf{D})^{1/2}$ and $\mathbf{u}^{\perp} = \mathbf{u}_2(\mathbf{D} \otimes \mathbf{D})^{1/2}$.

(b) $\|\mathbf{u}^{\parallel}\|_2 \leqslant \beta\|\mathbf{u}\|_2$ and $\|\mathbf{u}^{\perp}\|_2 \leqslant \beta\|\mathbf{u}\|_2$.

(c) If $\mathbf{u} \in \mathbb{R}^{n^2}$ represents a probability distribution (i.e., all entries of $\mathbf{u}$ are non-negative and $\|\mathbf{u}\|_1 = 1$), then $\mathbf{u}^{\parallel} = \boldsymbol{\pi} \otimes \boldsymbol{\pi}$.

We first look at the properties of the stochastic matrix $\left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right)$, which acts on the vector space $\mathbb{R}^{n^2}$ via right multiplication. The matrix and has the following properties:

1. Matrix $\left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right)$ fixes the vectors parallel to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$.
2. Matrix $\left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right)$ preserves the orthogonality to $\mathbf{1}$ (In fact, this is true for any stochastic matrix $\mathbf{M}$: $\langle \mathbf{u}\mathbf{M}, \mathbf{1}\rangle = \langle \mathbf{u}, \mathbf{1}\mathbf{M}^{\intercal}\rangle = \langle \mathbf{u}, \mathbf{1}\rangle = 0$ for $\mathbf{u} \perp \mathbf{1}$). Moreover, it shrinks vectors orthogonal to $\mathbf{1}$, because of the spectral expansion of G.

We will show that $\mathbf{Q}$ exhibits approximately the same properties.

Lemma 3.17. *For any vector $\mathbf{u} \in \mathbb{R}^{n^2}$ parallel to $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$, it holds that $\|(\mathbf{u}\mathbf{Q})^{\perp}\|_2 \leqslant \beta^4 n^{-1/2}\|\mathbf{u}\|_2$.*

Define the diagonal matrix $\mathbf{D}' \in \mathbb{R}^{n^2}$ such that

$$\mathbf{D'}_{(u,v)(u,v)} = \begin{cases} \delta \cdot \mathbf{D}_{u,u} & u = v, \\ (\mathbf{D} \otimes \mathbf{D})_{(u,v)(u,v)} & u \neq v. \end{cases}$$

Note that $\delta^2 \leqslant \mathbf{D}'_{(u,v)(u,v)} \leqslant (\mathbf{D} \otimes \mathbf{D})_{(u,v)(u,v)} \leqslant \Delta^2$ for all $u, v \in V[G]$.

Lemma 3.18. *For any vector $\mathbf{u} \in \mathbb{R}^{n^2}$ orthogonal to $\mathbf{1}$, it holds that $\mathbf{u}\mathbf{Q} \perp \mathbf{1}$ and*

$$\left\|\mathbf{u}\mathbf{Q}\mathbf{D}'^{-1/2}\right\|_2 \leqslant (1 - \alpha/8)\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2.$$

Given a distribution over $V[G] \times V[G]$, its two marginal distributions converge to the stationary distribution of $\mathbf{M}_G$ rapidly:

Lemma 3.19. *Let* $\mathbf{u}$ *represent a distribution over* $V[G] \times V[G]$*, and suppose matrix* $\mathbf{M}$ *is a convex combination of products of* $\mathbf{I} \otimes \mathbf{I}$*,* $\mathbf{I} \otimes \mathbf{M}_G$*,* $\mathbf{M}_G \otimes \mathbf{I}$ *and* $\mathbf{P}$*. Then for any* $k \in \mathbb{N}$ *it holds that*

$$\left\| \mathcal{T}_1(\mathbf{u}\mathbf{Q}^k\mathbf{M}) - \boldsymbol{\pi} \right\|_2 \leqslant (1-\alpha/2)^k \beta^{1/2},$$

and

$$\left\| \mathcal{T}_2(\mathbf{u}\mathbf{Q}^k\mathbf{M}) - \boldsymbol{\pi} \right\|_2 \leqslant (1-\alpha/2)^k \beta^{1/2}.$$

The joint distribution "almost" converges to the stationary distribution of $\mathbf{M}_G \otimes \mathbf{M}_G$ as well:

Lemma 3.20. *Let* $\mathbf{u}$ *represent a distribution over* $V[G] \times V[G]$*. Then for any* $k \in \mathbb{N}$ *it holds that*

$$\left\| \mathbf{u}\mathbf{Q}^k - \boldsymbol{\pi} \otimes \boldsymbol{\pi} \right\|_2 \leqslant 8\alpha^{-1}\beta^7 n^{-3/2} + (1-\alpha/8)^k \beta^2.$$

Lemma 3.21. *Let* $\mathbf{u}$ *represent a distribution over* $V[G] \times V[G]$*. And suppose matrix* $\mathbf{M}$ *satisfies*

$$\mathbf{M} = \left(\frac{\mathbf{M}_G + \mathbf{I}}{2} \otimes \mathbf{I} \right)^{k'} \quad \text{or} \quad \mathbf{M} = \left(\mathbf{I} \otimes \frac{\mathbf{M}_G + \mathbf{I}}{2} \right)^{k'}$$

for some $k' \in \mathbb{N}$*. Then for* $k \in \mathbb{N}$*, it holds that*

$$\begin{aligned} &\|\mathbf{u}\mathbf{Q}^k\mathbf{M} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 \\ &\leqslant (1-\alpha/2)^k \beta^{3/2} + (1-\alpha/2)^{k'} \cdot 8\alpha^{-1}\beta^8 n^{-3/2} + (1-\alpha/2)^{k'}(1-\alpha/8)^k \beta^3. \end{aligned}$$

Proof of Lemma 3.15. For $0 \leqslant k, k' \leqslant T/3$, define

$$\mathbf{U}_{k,k'} \triangleq \mathbf{E}_{S \in \mathcal{C}'_{T,k}, S' \in \mathcal{C}'_{T,k'}} \left[\mathbf{J}^{\star}_{S,S'} \right].$$

For the case $k \leqslant k'$, one can check that

$$\mathbf{U}_{k,k'} = \left(\frac{\mathbf{M}_G + \mathbf{I}}{2} \otimes \mathbf{I} \right)^{k'-k} \mathbf{Q}^{2T/3+k-k'} \left(\mathbf{I} \otimes \frac{\mathbf{M}_G + \mathbf{I}}{2} \right)^{k'-k}.$$

One way to see this is to note that $\mathbf{U}_{k,k'}$ corresponds to two colored random walks with color patterns that are randomly chosen as follows: (1) in the first k and last $T/3 - k'$ rounds, both two walks only take lazy steps; (2) from $(k+1)$th to k'th round, the first walk takes a lazy or non-lazy step with equal probability in each round whereas the second walk only takes lazy steps; (3) from $(k'+1)$th to $(2T/3+k)$th round, both two walks take a lazy or non-lazy step with equal probability in each round; (4) from $(2T/3+k+1)$th to $(2T/3+k')$th round, the first walk only takes lazy steps whereas the second walk takes a lazy or non-lazy step with equal probability in each round.

By Lemma 3.21 with $\mathbf{u} = \mathbf{e}_{(s,s)} \left(\frac{\mathbf{M}_G + \mathbf{I}}{2} \otimes \mathbf{I} \right)^{k'-k}$, we know that

$$\begin{aligned} \|\mathbf{e}_{(s,s)}\mathbf{U}_{k,k'} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 &\leqslant (1-\alpha/2)^{2T/3+k-k'} \beta^{3/2} + (1-\alpha/2)^{k'-k} \cdot 8\alpha^{-1}\beta^8 n^{-3/2} \\ &\quad + (1-\alpha/2)^{k'-k}(1-\alpha/8)^{2T/3+k-k'} \beta^3 \\ &\leqslant n^{-2} + (1-\alpha/2)^{k'-k} \cdot 8\alpha^{-1}\beta^8 n^{-3/2}, \end{aligned} \tag{3.4}$$

for sufficiently large $T = \Theta(\alpha^{-2} \log n)$. For the case $k \geqslant k'$, a symmetric argument gives the bound $n^{-2} + (1-\alpha/2)^{k-k'} \cdot 8\alpha^{-1}\beta^8 n^{-3/2}$.

Note that $\mathbf{x} = \mathbf{E}_{S,S' \in \mathcal{C}'_T} \left[\mathbf{e}_{(s,s)} \mathbf{J}^{\star}_{S,S'} \right] = \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'} \right]$, by (3.3) and the fact that sets $\mathcal{C}'_{T,k}$ have the same size for different k. Hence

$$\begin{aligned}
\|\mathbf{x} - \boldsymbol{\pi} \otimes \boldsymbol{\pi}\|_2 &= \left\| \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'} - \boldsymbol{\pi} \otimes \boldsymbol{\pi} \right] \right\|_2 \\
&\leqslant \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\left\| \mathbf{e}_{(s,s)} \mathbf{U}_{k,k'} - \boldsymbol{\pi} \otimes \boldsymbol{\pi} \right\|_2 \right] \\
&\leqslant \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[n^{-2} + (1-\alpha/2)^{|k'-k|} \cdot 8\alpha^{-1}\beta^8 n^{-3/2} \right] \\
&\leqslant n^{-2} + 8\alpha^{-1}\beta^8 n^{-3/2} \cdot \mathbf{E}_{0 \leqslant t \leqslant T/6} \left[(1-\alpha/2)^t \right] \\
&\leqslant n^{-2} + 96\alpha^{-2}\beta^8 n^{-3/2} T^{-1} = o(n^{-3/2}),
\end{aligned}$$

where the second inequality follows from (3.4) and the last step follows from $T = \Theta(\alpha^{-2} \log n)$.

Next we prove the second claim. Assume $0 \leqslant k \leqslant k' \leqslant T/3$. By Lemma 3.19 with $\mathbf{u} = \mathbf{e}_{(s,s)} \left(\frac{\mathbf{M}_G + \mathbf{I}}{2} \otimes \mathbf{I} \right)^{k'-k}$, we have

$$\left\| \mathcal{T}_1(\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'}) - \boldsymbol{\pi} \right\|_2 \leqslant (1-\alpha/2)^{2T/3+k-k'} \beta^{1/2} \leqslant (1-\alpha/2)^{T/3} \beta^{1/2}.$$

For the case $k' \leqslant k$ a symmetric argument gives the same bound.

Also note that $\mathbf{y} = \mathcal{T}_1(\mathbf{x}) = \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\mathcal{T}_1(\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'}) \right]$. Hence

$$\begin{aligned}
\|\mathbf{y} - \boldsymbol{\pi}\|_2 &= \left\| \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\mathcal{T}_1(\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'}) - \boldsymbol{\pi} \right] \right\|_2 \\
&\leqslant \mathbf{E}_{0 \leqslant k,k' \leqslant T/3} \left[\left\| \mathcal{T}_1(\mathbf{e}_{(s,s)} \mathbf{U}_{k,k'}) - \boldsymbol{\pi} \right\|_2 \right] \\
&\leqslant (1-\alpha/2)^{T/3} \beta^{1/2} = O(n^{-2})
\end{aligned}$$

for sufficiently large $T = \Theta(\alpha^{-2} \log n)$. ■

3.3 Protocol for Dense Graphs

The second statement of Theorem 3.3 is based on Theorem 3.10, and uses another idea that once $\Omega(n)$ nodes are informed in an expander graph with $\delta = n$, we can just look at the rumors sent by this fixed set of nodes and ignore the others without impairing the rumor spreading rate too much. Our construction is based on expander walks and pairwise independent generators. The protocol is as follows.

Protocol 3 (Protocol for Dense Graphs)**.** *Let $m = n^{\Theta(1)}$ be a sufficiently large power of 2. Pick the following objects:*

1. *An explicit pairwise independent generator $\mathcal{G} = (\mathcal{G}_0, \ldots, \mathcal{G}_{n-1}) : \{0,1\}^\ell \to [m]^n$, where $\ell = O(\log n)$.*
2. *An explicit degree-D expander graph G' with spectral gap γ and node set $\{0,1\}^\ell$ where D and γ are positive constants.*

The initial node having the rumor chooses a random string x as in Protocol 1. It also picks random strings $y \in \{0,1\}^\ell$ and $z = (z_1, \ldots, z_T) \in [D]^T$, where $T = \Theta(\alpha^{-2} \cdot \log n)$. Once one node gets the rumor, it gets the ID u. In first T rounds, each node sends the rumor using the random string x as in Protocol 1, except that (x, y, z) instead of x is appended with the rumor. In the $(T+i)$th round where $1 \leqslant i \leqslant T$, node u chooses a neighbor v with index $\mathcal{G}_u(y) \bmod \deg(u)$ in its adjacency list. It then replaces y with y', the z_ith neighbor of y in G'. The triple of random strings (x, y', z) is then appended with the rumor and sent to the node v.

Proposition 3.22. *Protocol 3 uses $O(\log n(\log\log n + \log(1/\alpha)))$ random bits in total.*

Now we prove the second statement of Theorem 3.3.

Proof of the second statement of Theorem 3.3. By Theorem 3.10, with high probability, all but $1 - o(1)$ nodes are informed after $O(\alpha^{-2} \cdot \log n)$ rounds. Condition on this event which only depends on x. Then y and z are still uniformly distributed. We will show that all nodes are informed in the next T rounds with probability at least $1 - 1/n$. Let $y_0 = y$ and y_i be the z_ith neighbor of y_{i-1} in the expander graph G' picked in Protocol 3, i.e., $(y_0, \ldots, y_T)$ is an expander walk on G' with edge labels $z_1, \ldots, z_T$. Note that in the $(T+i)$th round, all nodes use y_i as the seed of the pairwise independent generator $\mathcal{G}$.

Fix an uninformed node v. Let I be the set of neighbors of v that are informed in the first T rounds. Then $|I| \geqslant \deg(v) - o(n) = \Theta(\deg(u))$. Consider a random $w \in \{0,1\}^\ell$. For $u \in I$, let X_u be the boolean random variable whose value is 1 iff u informs v using seed w, i.e., $\mathcal{G}_u(w) \bmod \deg(u)$ equals the index of v in the adjacency list of u. We have $|\mathbf{E}[X_u] - 1/\deg(u)| \leqslant 2/m$ for any $u \in I$ and $|\mathbf{E}[X_u X_{u'}] - \mathbf{E}[X_u]\mathbf{E}[X_{u'}]| \leqslant 2/m$ for any distinct $u, u' \in I$, by the fact that $\mathcal{G}$ is a pairwise independent generator, together with Lemma 2.2. By Cauchy-Schwarz inequality, we have

$$\mathbf{Pr}\left[\sum_{u\in I} X_u > 0\right] = \mathbf{E}\left[\mathbf{1}_{\sum_{u\in I} X_u > 0}\right] \geqslant \frac{\left(\mathbf{E}\left[\sum_{u\in I} X_u\right]\right)^2}{\mathbf{E}\left[\left(\sum_{u\in I} X_u\right)^2\right]}$$

where

$$\mathbf{E}\left[\sum_{u\in I} X_u\right] = \sum_{u\in I} \mathbf{E}[X_u] \geqslant \sum_{u\in I}\left(\frac{1}{\deg(u)} - \frac{2}{m}\right) = \Omega(|I|/\deg(u)) = \Omega(1).$$

and

$$\begin{aligned}
\mathbf{E}\left[\left(\sum_{u\in I} X_u\right)^2\right] &= \sum_{u\in I} \mathbf{E}[X_u] + \sum_{\substack{u,u'\in I\\ u\neq u'}} \mathbf{E}[X_u X_{u'}] \\
&\leqslant \sum_{u\in I}\left(\frac{1}{\deg(u)} + \frac{2}{m}\right) + \sum_{\substack{u,u'\in I\\ u\neq v}}\left(\frac{1}{\deg(u)\deg(u')} + \frac{2}{m}\right) \\
&= O(|I|/\deg(u) + |I|^2/(\deg(u)\deg(u'))) = O(1).
\end{aligned}$$

So $\mathbf{Pr}\left[\sum_{u\in I} X_u > 0\right] \geqslant c$ for some constant $c > 0$. Let S be the set of seeds $w \in \{0,1\}^\ell$ such that there exists some node $u \in I$ that informs v using seed w. Then $|S| \geqslant c \cdot 2^\ell$. By the hitting property of expander walks [1], we have

$$\mathbf{Pr}_{y,z}[\, y_i \notin S :\ \forall 0 \leqslant i \leqslant T\,] \leqslant 2^{-\Omega(T)}.$$

For sufficiently large $T = \Theta(\alpha^{-2} \cdot \log n)$, this probability is at most $1/n^2$. So for any fixed $v \in V[G] \setminus I$, the probability that v is not informed in T rounds is at most $1/n^2$. The claim follows by the union bound. ∎

4 Protocol for General Graphs

We further study the protocol for general graphs. Formally let G be any graph with n nodes, $\Delta(G)/\delta(G) = O(1)$ and conductance ϕ, where

$$\phi(G) \triangleq \min_{S\subseteq V, 0<|S|<n} \frac{e(S, V\setminus S)}{\min\{\mathrm{vol}(S), \mathrm{vol}(V\setminus S)\}}.$$

Theorem 4.1. *Let G be any graph with $\Delta/\delta = O(1)$, and conductance ϕ. Then there is a protocol using $O\big((1/\phi)\cdot\log n\cdot(\log\log n + \log\Delta)\big)$ random bits in total, so that with high probability every node in G gets informed in $O((1/\phi)\log n)$ rounds.*

4.1 Protocol

Our protocol is based on pairwise independent generators and unbalanced expanders with near-optimal expansion. Here different rounds use different random bits. In contrast to $O(n \log n)$ random bits per round used in the truly random protocol and $\Omega(\log n)$ random bits per round used in [24], we show that $O(\log \log n + \log \Delta)$ random bits per round suffice to spread the rumor efficiently on general graphs G with $\Delta(G)/\delta(G) = O(1)$. The formal description of our protocol is as follows:

Protocol 4 (Protocol for General Graphs)**.** *Let $\varepsilon = \Delta^{-\Theta(1)}$ be sufficiently small and $m = 2^{\lceil \log(4/\varepsilon) \rceil}$. Pick the following objects:*

- *An explicit $(K, (1-\varepsilon^2/4)D)$-expander $\Gamma : [n^c] \times [D] \to \bigsqcup_{i \in [D]} [M_i]$, where $K = 2$, $D = \left(\frac{\log n}{\varepsilon}\right)^{O(1)}$ and $M_0 = \cdots = M_{D-1} = M \leqslant D$.*
- *An explicit pairwise independent generator $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_M) : \{0,1\}^\ell \to [m]^M$, where $\ell = O(\log m + \log M) = O(\log \log n + \log \Delta)$.*

These two objects $\mathcal{G}$ and Γ can be uniquely constructed from n^c and $\Delta^{\Theta(1)}$, and hence are known to every informed node.

The initial node having the rumor chooses a random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, y_i) \in [D] \times \{0,1\}^\ell$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, node u computes $r = \Gamma(u, x_i)$ that is in $[M_u]$, the uth copy of $[M]$. It then chooses the neighbor with index $\mathcal{G}_r(y_i) \bmod \deg(u)$ in its adjacency list to send the rumor.

The protocol above presents a nice "two-level hashing" framework: The first level is based on a pairwise independent generator $\mathcal{G}$. While the PRG-based protocol in [24] needs to generate $O(n)$ blocks and different nodes need to use different blocks, our protocol only needs $M = (\Delta \log n)^{O(1)}$ blocks and hence $O(\log \log n + \log \Delta)$ random bits suffice for this purpose. The second level uses unbalanced expanders to map node ID $u \in [n^c]$ to $r \in \left[\Delta^{O(1)}\right]$ by using $\log \log n + \log \Delta$ random bits. After these, node u uses the value of the rth block of $\mathcal{G}$ to choose the neighbors. It is easy to see that every informed node u only needs $O(\operatorname{poly} \log n)$ arithmetic operations per round in order to determine its neighbor.

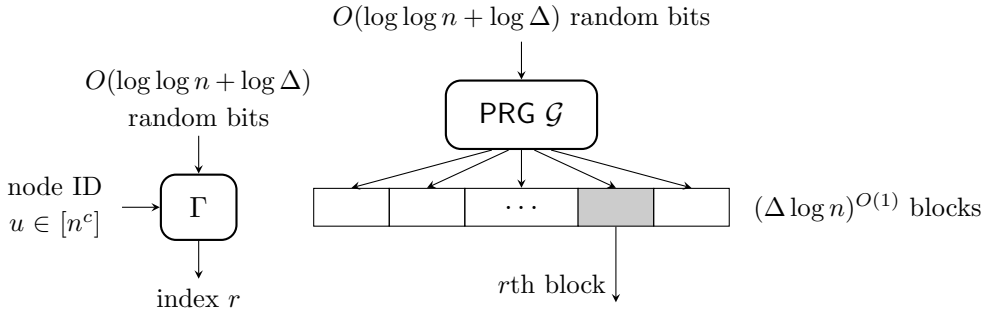

Figure 1: Illustration of the protocol for general graphs. Every node u uses an unbalanced expander Γ to generate an index r, and uses the rth block of PRG $\mathcal{G}$ to choose a neighbor to send the rumor.

Proposition 4.2. *Assume that Protocol 4 finishes in T rounds. Then it uses $O(T \cdot (\log \log n + \log \Delta))$ random bits in total.*

Remark 4.3. *Using the explicit constructions of unbalanced expanders in [26] and pairwise independent generators in [6], our protocol is very simple and can be described as follows: Let*

$\mathbb{F}_q$ be a finite field of size $q = (\Delta \log n)^{\Theta(1)}$ for a sufficiently large q. Let $E(X)$ be an irreducible polynomial of degree m over $\mathbb{F}_q$ such that $n^c \leqslant q^m = n^{O(1)}$. We embed $[n^c]$ into the finite field $\mathbb{F}_q^m \cong \mathbb{F}_q[X]/(E(X))$. Through this we can identify every node with ID $u \in [n^c]$ with a polynomial p_u of degree at most $m-1$ over $\mathbb{F}_q$. The protocol then uses the random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, y_i) = (x_i, a_i, b_i) \in \mathbb{F}_q^3$. Then node u computes $z = a_i \cdot p_u(x_i) + b_i$ (over $\mathbb{F}_q$) in the ith round, and chooses the neighbor with index $(z \bmod \deg(u))$ in its adjacency list to send the rumor.

4.2 Analysis

We start by analyzing a single round t and see the properties of our protocol. Let I_t be the set of informed nodes after round t, and U_t the set of uninformed nodes after round t. Remember that all the random choices in round t are determined by (x_t, y_t).

We need the following lemma:

Lemma 4.4. *Fix any round $0 \leqslant t < T$. For any $u \in U_t$, $v \in I_t$, let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$. Then it holds that*

1. *$|\mathbf{E}[X_{v\to u}] - 1/\deg(v)| \leqslant \varepsilon$ for any $u \in U_t$, $v \in I_t$,*
2. *$\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant \varepsilon$ for any $u, u' \in U_t$, $v, v' \in I_t$, $(u, v) \neq (u', v')$.*

Proof. For any $u \in U_t$, $v \in I_t$, suppose the index of u in the adjacency list of v is z. By construction, $X_{v\to u}$ equals 1 iff $\mathcal{G}_{\Gamma(v,x_t)}(y_t) \bmod \deg(v) = z$. Fix x_t. The fact that $\mathcal{G}$ is a pairwise independent generator together with Lemma 2.2 shows that $|\mathbf{E}[X_{v\to u}] - 1/\deg(v)| \leqslant 2/m \leqslant \varepsilon$.

For any $u, u' \in U_t$, $v, v' \in I_t$, first assume $v \neq v'$. Suppose the index of u (resp. u') in the adjacency list of v (resp. v') is z (resp z'). By construction, $X_{v\to u}$ equals 1 iff $\mathcal{G}_{\Gamma(v,x_t)}(y_t) \bmod \deg(v) = z$, and similarly for $X_{v'\to u'}$. By Lemma 2.13 and the fact that Γ is a $(K, (1-\varepsilon^2/4)D)$-expander, the event $|\{\Gamma(v, x_t), \Gamma(v', x_t)\}| \geqslant (1-\varepsilon/2)\cdot 2 > 1$ occurs with probability at least $1 - \varepsilon/2$ over the choices of x_t. Condition on any x_t such that this event occurs. We have $\Gamma(v, x_t) \neq \Gamma(v', x_t)$. Using the fact that $\mathcal{G}$ is pairwise independent together with Lemma 2.2, we have $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant 2/m$. For the other choices of x_t, we have $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant 1$ since $X_{v\to u}, X_{v'\to u'}$ are boolean. Therefore $\mathbf{Cov}[X_{v\to u}, X_{v'\to u'}] \leqslant (1-\varepsilon/2)(2/m) + (\varepsilon/2) \leqslant \varepsilon$ for random x_t.

Now assume $v = v'$ and hence $u \neq u'$. We have

$$\begin{aligned}\mathbf{Cov}[X_{v\to u}, X_{v\to u'}] &= \mathbf{E}[X_{v\to u} \cdot X_{v\to u'}] - \mathbf{E}[X_{v\to u}] \cdot \mathbf{E}[X_{v\to u'}] \\ &= 0 - \mathbf{E}[X_{v\to u}] \cdot \mathbf{E}[X_{v\to u'}] \leqslant 0.\end{aligned}$$

■

Next we prove the following lemma:

Lemma 4.5. *Fix a round $0 \leqslant t < T$ and the set I_t of informed nodes before round $t+1$. Fix also an arbitrary set of edges $F \subseteq E(I_t, U_t)$. Let J be the set of nodes that become informed in round $t+1$ if we consider only transmissions of the rumor along the edges in F.*

1. *$\mathbf{Pr}[J \neq \emptyset] \geqslant c_1 \min\{|F|/\Delta, 1\}$ for some constant $c_1 > 0$.*
2. *If $|F| = \Omega(\Delta)$ then $\mathbf{Pr}[|J| \geqslant c_2|F|/\Delta] \geqslant c_3$ for some constant $c_2, c_3 > 0$.*

Proof. Let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$.

We first prove (1). Let $k = |F|$ and suppose $F = \{(v_0, u_0), \ldots, (v_{k-1}, u_{k-1})\}$. Let $X = \sum_{i\in[k]} X_{v_i\to u_i}$. Then by Cauchy-Schwarz inequality, $\mathbf{E}[\mathbf{1}_{X>0}] \geqslant (\mathbf{E}[X])^2/\mathbf{E}[X^2]$. By Lemma 4.4,

$$\mathbf{E}[X] = \sum_{i\in[k]} \mathbf{E}[X_{v_i\to u_i}] \geqslant k(1/\Delta - \varepsilon) = \Omega(|F|/\Delta)$$

and

$$\begin{aligned}\mathbf{E}\left[X^2\right] &= \sum_{i,j\in[k]} \mathbf{E}\left[X_{v_i\to u_i}X_{v_j\to u_j}\right] \\ &= \sum_{i\in[k]} \mathbf{E}\left[X_{v_i\to u_i}\right] + \sum_{\substack{i,j\in[k]\\ i\neq j}} \left(\mathbf{E}\left[X_{v_i\to u_i}\right]\mathbf{E}\left[X_{v_j\to u_j}\right] + \mathbf{Cov}\left[X_{v_i\to u_i}, X_{v_j\to u_j}\right]\right) \\ &\leqslant k(1/\delta+\varepsilon) + (k^2-k)((1/\delta+\varepsilon)^2+\varepsilon) = O(|F|/\Delta + |F|^2/\Delta^2)\end{aligned}$$

where we use the conditions that $\Delta/\delta = O(1)$ and $\varepsilon = \Delta^{-\Theta(1)}$ is sufficiently small. So

$$\mathbf{Pr}\left[J\neq\emptyset\right] = \mathbf{E}\left[\mathbf{1}_{X>0}\right] \geqslant (\mathbf{E}\left[X\right])^2/\mathbf{E}\left[X^2\right] = \Omega(\min\{|F|/\Delta, 1\})$$

and the first statement follows.

Next we prove the second statement. For $u\in U_t$, let F_u be the set of edges in F incident to u, Z_u be the boolean random variable whose value is 1 iff u is informed in round $t+1$ via edges in F_u, and $X_u = \sum_{(v,u)\in F_u} X_{v\to u}$. So $Z_u = \mathbf{1}_{X_u>0}$ and $|J| = \sum_{u\in U_t} Z_u$. For $u\in U_t$, $\mathbf{E}\left[Z_u\right] = \mathbf{E}\left[\mathbf{1}_{X_u>0}\right] \geqslant (\mathbf{E}\left[X_u\right])^2/\mathbf{E}\left[X_u^2\right] = \Omega(|F_u|/\Delta)$ by a similar argument as above. So $\mathbf{E}\left[|J|\right] = \Omega(\sum_{u\in U_t}|F_u|/\Delta) = \Omega(|F|/\Delta)$. Suppose $\mathbf{E}\left[|J|\right] \geqslant c|F|/\Delta$ for constant $c>0$.

On the other hand, for any $c_2 \geqslant 0$, we have

$$\begin{aligned}\mathbf{E}\left[|J|\right] &= \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] + \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\cdot|J|\right] \\ &\leqslant \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] + \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\right]\cdot c_2|F|/\Delta\end{aligned}$$

and hence $\mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right] \geqslant \mathbf{E}\left[|J|\right] - \mathbf{E}\left[\mathbf{1}_{|J|<c_2|F|/\Delta}\right]\cdot c_2|F|/\Delta \geqslant (c-c_2)|F|/\Delta$. Pick $c_2 = c/2$. By Cauchy-Schwarz inequalty, we have

$$\mathbf{Pr}\left[|J|\geqslant c_2|F|/\Delta\right] = \mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\right] \geqslant \frac{\left(\mathbf{E}\left[\mathbf{1}_{|J|\geqslant c_2|F|/\Delta}\cdot|J|\right]\right)^2}{\mathbf{E}\left[|J|^2\right]} \geqslant \frac{((c-c_2)|F|/\Delta)^2}{\mathbf{E}\left[|J|^2\right]}. \quad (4.1)$$

Note that

$$\begin{aligned}\mathbf{E}\left[|J|^2\right] &= \sum_{u\in U_t}\mathbf{E}\left[Z_u\right] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\mathbf{E}\left[Z_uZ_{u'}\right] \\ &\leqslant \mathbf{E}\left[|J|\right] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\mathbf{E}\left[X_uX_{u'}\right] \\ &= \mathbf{E}\left[|J|\right] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\sum_{\substack{(v,u)\in F_u\\ (v',u')\in F_{u'}}}\left(\mathbf{E}\left[X_{v\to u}\right]\mathbf{E}\left[X_{v'\to u'}\right] + \mathbf{Cov}\left[X_{v\to u}, X_{v'\to u'}\right]\right) \\ &\leqslant \mathbf{E}\left[|J|\right] + \sum_{\substack{u,u'\in U_t\\ u\neq u'}}\left(\left(\sum_{(v,u)\in F_u}\mathbf{E}\left[X_{v\to u}\right]\right)\left(\sum_{(v',u')\in F_{u'}}\mathbf{E}\left[X_{v'\to u'}\right]\right) + |F_u||F_{u'}|\varepsilon\right) \\ &= \mathbf{E}\left[|J|\right] + O\left(\sum_{\substack{u,u'\in U_t\\ u\neq u'}}|F_u||F_{u'}|/\delta^2\right) \\ &= \mathbf{E}\left[|J|\right] + O\left(\left(\sum_{u\in U_t}|F_u|\right)^2/\delta^2\right) \\ &= \mathbf{E}\left[|J|\right] + O\left(|F|^2/\delta^2\right).\end{aligned}$$

Here $\mathbf{E}[\,|J|\,] = \sum_{u\in U_t}\mathbf{E}[\,Z_u\,] \leqslant \sum_{u\in U_t}\mathbf{E}[\,X_u\,] = \sum_{u\in U_t} O(|F_u|/\delta) = O(|F|/\delta)$. Using the condition $|F| = \Omega(\Delta) = \Omega(\delta)$, we have $\mathbf{E}\left[\,|J|^2\,\right] = O\left(|F|^2/\delta^2\right)$. Substitute it in (4.1) and use the condition $\Delta/\delta = O(1)$, and then (2) follows. ■

From here our analysis follows [24].

Proof of Proposition 4.2. The proof is divided into four phases, depending on the number of informed nodes $|I_t|$.

Phase 1: $1 \leqslant |I_t| \leqslant 1/\phi$. This phase is divided into several subphases. For every $1 \leqslant i \leqslant \log(1/\phi)$, subphase i begins when the number of informed nodes is at least 2^{i-1} and ends when this number is at least 2^i. Assume that we are at the beginning of the ith subphase. Fix an arbitrary round t of the ith subphase and the set of informed nodes I_t; thus, $2^{i-1} \leqslant |I_t| < 2^i$. We consider the number of nodes that become informed in round $t+1$. Applying Lemma 4.5(1) with $F = E(I_t, U_t)$ gives

$$\mathbf{Pr}\,[\,|I_{t+1} \setminus I_t| \geqslant 1\,] \geqslant c_1 \min\{e(I_t,U_t)/\Delta, 1\} \geqslant c_1 \min\{\phi(\delta/\Delta)|I_t|, 1\},$$

Let $p \triangleq c_1 \min\{\phi(\delta/\Delta)|I_t|, 1\}$. We have $p = O(\phi|I_t|)$ since $|I_t| \leqslant 1/\phi$ and $\Delta/\delta = \Theta(1)$. Therefore, the expected time to increase $|I_t|$ from 2^{i-1} to 2^i is at most $2^{i-1}/p = O(1/\phi)$. By Markov's inequality,

$$\mathbf{Pr}\left[\,|I_{t+\tau}| \leqslant 2^i \;\middle|\; |I_t| \geqslant 2^{i-1}\,\right] \leqslant 1/2$$

for some $\tau = O(\phi^{-1})$. Hence the time to complete Phase 1 can be upper bounded by $\tau = O((1/\phi))$ multiplied with the sum of $\log(1/\phi) = O(\log n)$ independent geometric random variables each with parameter $1/2$. Applying a Chernoff bound for the sum of independent geometric random variables yields that the number of rounds required for Phase 1 is at most $O((1/\phi)\cdot\log n)$ with high probability.

Phase 2: $1/\phi \leqslant |I_t| \leqslant n/2$**.** Fix a round t and the set of informed nodes I_t. We apply Lemma 4.5(2), with $F = E(I_t, U_t)$. Note that the precondition $|F| = \Omega(\Delta)$ is satisfied, as

$$|F| = e(I_t, U_t) \geqslant \phi \cdot \delta \cdot |I_t| \geqslant \phi \cdot \delta \cdot (\Delta/\delta) \cdot (1/\phi) = \Omega(\Delta).$$

Hence we conclude from Lemma 4.5(2) that

$$\mathbf{Pr}\,[\,|I_{t+1} \setminus I_t| \geqslant c_2 \cdot \phi \cdot \delta \cdot |I_t|/\Delta\,] \geqslant c_3,$$

for some constant $c_2, c_3 > 0$. When this event occurs, we have $|I_{t+1}| \geqslant (1 + c_2\phi\delta/\Delta)|I_t|$. So, the number of rounds until we have $|I_t| \leqslant n/2$ can be upper bounded by the sum of $\log_{1+c_2\phi\delta/\Delta}(n/2) = O((1/\phi)\log n)$ independent geometric random variables with parameters c_3. Using again the Chernoff bound we obtain that Phase 2 is completed within at most $O((1/\phi)\log n)$ rounds with high probability.

Phase 3: $n/2 \leqslant |I_t| \leqslant n - 1/\phi$**.** The analysis is the same as in Phase 2 with the roles of I_t and U_t switched.

Phase 4: $n - 1/\phi \leqslant |I_t| \leqslant n$**.** Again, the analysis is the same as in Phase 1 with the roles of I_t and U_t switched.

Since each of the four phases requires only $O((1/\phi)\cdot\log n)$ rounds with high probability, the result follows by applying the union bound. ■

5 Protocol For Strong Expander Graphs

We further study the protocols for strong expander graphs. Let $\mathcal{L}$ be the normalized Laplacian matrix of G defined by

$$\mathcal{L}_{u,v} = \begin{cases} 1 & \text{if } u = v \text{ and } \deg(u) \neq 0, \\ -\dfrac{1}{\sqrt{\deg(u) \cdot \deg(v)}} & \text{if } u \text{ and } v \text{ are adjacent}, \\ 0 & \text{otherwise.} \end{cases}$$

i.e. $\mathcal{L} \triangleq \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}_G\mathbf{D}^{-1/2}$. Let the eigenvalues of $\mathcal{L}$ be $0 = \lambda_0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{n-1}$. Moreover define $\lambda \triangleq \max\{1 - \lambda_1, \lambda_{n-1} - 1\}$. We call a family of graphs $\{G_i\}_{i \geqslant 0}$ *strong expander graphs* if $\Delta/\delta = 1 + o(1)$ and $\lambda = o(1)$ for any G_i. This graph family includes several interesting graphs, e.g. Ramanujan graphs, complete graphs, random graphs $G(n, p)$ with $p = \omega(\log n/n)$, and random d-regular graph where d is any increasing function of n. Our result for strong expander graphs is as follows:

Theorem 5.1. *Let G be a graph such that $\Delta/\delta = 1 + o(1)$ and $\lambda = o(1)$. Then there is a protocol using $O(\log n \cdot (\log\log n + \log\Delta))$ random bits in total, so that with high probability every node in G gets informed in $\log n + \ln n + o(\log n)$ rounds.*

The runtime in Theorem 5.1 matches the precise runtime for the truly random protocol [15, 16, 18], and is known to be tight [16]. Moreover, our protocol uses $O(\log n \cdot (\log\log n + \log\Delta))$ random bits in total, in contrast to $O(n \log^2 n)$ random bits in the truly random protocol, and $O(\log^3 n)$ random bits in the previous best result [24]. For complete graphs, we use $O(\log^2 n)$ random bits in total, in contrast to $O(n \log n)$ random bits used in the quasirandom rumor spreading protocol [19].

5.1 Protocol

Our protocol follows the framework of the pseudorandom protocol in [24]. However, instead of using Nisan's generators, we use PRGs for combinatorial rectangles and pairwise independent generators, together with unbalanced expanders. The construction of our protocol implies that, in order to maintain the precise rumor spreading time as in the truly random protocol, we only need $O(\log\log n + \log\Delta)$ random bits per round, in contrast to $O(\log^2 n)$ random bits in [24].

The formal description of our protocol is as follows:

Protocol 5 (Protocol for Strong Expander Graphs)**.** *Let $\varepsilon = \Delta^{-\Theta(1)}$ be sufficiently small, $\varepsilon' = 2^{-\sqrt{\log\log n}}$, and $m = \Theta((\log n)/\varepsilon)$ a power of 2. Pick the following objects:*

- *An explicit $(\leqslant K, (1 - \varepsilon^2/4)D)$-expander $\Gamma : [n^c] \times [D] \to \bigsqcup_{i \in [D]}[M_i]$, where $K = \Delta$, $D = \left(\frac{\log n}{\varepsilon}\right)^{O(1)}$ and $M_0 = \cdots = M_{D-1} = M \leqslant \max\{D, \Delta^{O(1)}\}$.*
- *An explicit function $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_M) : \{0,1\}^\ell \to [m]^M$ that is both a pairwise independent generator and a PRG ε'-fooling $\mathcal{M}_{[m]^M, M}$, where $\ell = O(\log m + \log M + \log(1/\varepsilon')\log\log(1/\varepsilon')\log\log\log(1/\varepsilon')) = O(\log\log n + \log\Delta)$.*

These two objects $\mathcal{G}$ and Γ can be uniquely constructed from n^c and $\Delta^{\Theta(1)}$, and hence are known to every informed node.

The initial node having the rumor chooses a random string $(s_1, \ldots, s_T)$ where every s_i is of the form $(x_i, y_i) \in [D] \times \{0,1\}^\ell$. This random string is appended with the rumor and sent to other nodes. Once one node gets the rumor, it gets the ID u. In the ith round, node u computes $r = \Gamma(u, x_i)$ that is in $[M_u]$, the uth copy of $[M]$. It then chooses the neighbor with index $\mathcal{G}_r(y_i) \bmod \deg(u)$ in its adjacency list to send the rumor.

Proposition 5.2. *Assume that Protocol 5 finishes in T rounds. Then it uses $O(T \cdot (\log\log n + \log\Delta))$ random bits in total.*

5.2 Analysis

To relate the spectral expansion of G with the expansion property, we use the following expander mixing lemma for general graphs.

Lemma 5.3 (Expander Mixing Lemma for General Graphs [10])**.** *Let G be a general graph. Then for any subset X and Y it holds that*

$$\left| e(X,Y) - \frac{\mathrm{vol}(X)\cdot\mathrm{vol}(Y)}{\mathrm{vol}(G)} \right| \leqslant \lambda \cdot \frac{\sqrt{\mathrm{vol}(X)\cdot\mathrm{vol}(Y)\cdot\mathrm{vol}(\overline{X})\cdot\mathrm{vol}(\overline{Y})}}{\mathrm{vol}(G)}.$$

In order to prove Theorem 5.1, it suffices to show the following lemma:

Lemma 5.4. *Let G be a graph that satisfies the preconditions of Theorem 5.1. Then with high probability all the following statements hold:*

- ***Phase I*** *Suppose $1 \leqslant |I_t| \leqslant n/\log n$. Then there is $\tau = \log n + o(\log n)$ such that $|I_{t+\tau}| > n/\log n$.*
- ***Phase II*** *Suppose $n/\log n \leqslant |I_t| \leqslant n - n/\log n$. Then there is $\tau = o(\log n)$ such that $|I_{t+\tau}| > n - n/\log n$.*
- ***Phase III*** *Suppose $|I_t| \geqslant n - n/\log n$. Then there is $\tau = \ln n + o(\log n)$ such that $|I_{t+\tau}| = n$.*

Proof. For any round t and $u \in U_t$, $v \in I_t$, let $X_{v\to u}$ be the boolean random variable whose value is 1 iff v informs u in round $t+1$. Note that Γ is a $(\leqslant K, (1-\varepsilon^2/4)D)$-expander and hence a $(2, (1-\varepsilon^2/4)D)$-expander. And $\mathcal{G}$ is a pairwise independent generator. Then we observe that the statements in Lemma 4.4 hold here as well by the same proof.

Phase I. By Lemma 5.3 we have

$$\begin{aligned} e(I_t, U_t) &\geqslant \frac{\mathrm{vol}(I_t)\cdot\mathrm{vol}(U_t)}{\mathrm{vol}(G)} - \lambda\cdot\frac{\mathrm{vol}(I_t)\cdot\mathrm{vol}(U_t)}{\mathrm{vol}(G)} \\ &\geqslant (1-\lambda)\cdot\frac{\mathrm{vol}(I_t)\cdot(\mathrm{vol}(G)-\mathrm{vol}(I_t))}{\mathrm{vol}(G)} \\ &\geqslant (1-\lambda)\cdot\delta\cdot|I_t|\left(1-\frac{\Delta\cdot|I_t|}{nd}\right). \end{aligned}$$

Since $\lambda = o(1)$ and $|I_t| \leqslant n/\log n$, we have

$$\begin{aligned} e(I_t, U_t) &\geqslant (1-o(1))\cdot\Delta\cdot|I_t|\left(\frac{\delta}{\Delta} - \frac{\delta}{d\cdot\log n}\right) \\ &\geqslant \left(1 - \frac{1}{\log n} - o(1)\right)\cdot\Delta\cdot|I_t|. \end{aligned} \tag{5.1}$$

Hence

$$|N(I_t)\setminus I_t| \geqslant \frac{e(I_t,U_t)}{\Delta} \geqslant \left(1-\frac{1}{\log n} - o(1)\right)\cdot|I_t|.$$

Define $\gamma \triangleq \lambda + \frac{1}{\log n}$, and $A \triangleq \{u \in N(I_t)\setminus I_t : |N(u)\cap I_t| \geqslant 2d\sqrt{\gamma}\}$. Then $e(A, I_t) \geqslant |A|\cdot 2d\cdot\sqrt{\gamma}$. On the other hand by Lemma 5.3 it holds that

$$\begin{aligned} e(A, I_t) &\leqslant \frac{\mathrm{vol}(A)\cdot\mathrm{vol}(I_t)}{\mathrm{vol}(G)} + \lambda\sqrt{\mathrm{vol}(A)\cdot\mathrm{vol}(I_t)} \\ &\leqslant \frac{\Delta^2\cdot|A|\cdot|I_t|}{nd} + \gamma\Delta\cdot\sqrt{|A|\cdot|I_t|}. \end{aligned}$$

By the definition of set A we have $e(A, I_t) \geqslant 2d\sqrt{\gamma} \cdot |A|$, and hence

$$\begin{aligned} |A| \cdot 2d \cdot \sqrt{\gamma} &\leqslant \frac{\Delta^2 \cdot |A| \cdot |I_t|}{nd} + \gamma\Delta \cdot \sqrt{|A| \cdot |I_t|} \\ &\leqslant (1 + o(1)) \cdot \frac{\Delta \cdot |A|}{\log n} + \gamma\Delta \cdot \sqrt{|A| \cdot |I_t|}, \end{aligned}$$

which implies $|A| \leqslant \gamma \cdot |I_t|$.

Now define $B \triangleq N(I_t) \setminus I_t \setminus A$. We have

$$e(B, I_t) = e(N(I_t), I_t) - e(A, I_t) \geqslant \left(1 - \frac{1}{\log n} - o(1) - \gamma\right) \Delta \cdot |I_t|.$$

With the above estimate at hand, we compute the expected value of $|I_t \cap B|$. Note that for any $u \in B$, the chance that it gets informed in round $t + 1$ is

$$p_{t+1}(u) \triangleq \mathbf{Pr}\left[\bigvee_{v \in N(u) \cap I_t} (X_{v \to u} = 1)\right]$$

which is lower bounded by

$$\sum_{v \in N(u) \cap I_t} \mathbf{Pr}\left[\, X_{v \to u} = 1 \,\right] - \sum_{\substack{v_1, v_2 \in N(u) \cap I_t \\ v_1 < v_2}} \mathbf{Pr}\left[\bigwedge_{i=1,2} (X_{v_i \to u} = 1)\right]$$

by Bonferroni inequalities. Hence

$$\begin{aligned} p_{t+1}(u) &\geqslant |N(u) \cap I_t| \left(\frac{1}{\Delta} - \varepsilon\right) - \binom{|N(u) \cap I_t|}{2} \left(\frac{1}{\delta^2} + \varepsilon\right) \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} - (1 + o(1)) \cdot \binom{|N(u) \cap I_t|}{2} \cdot \frac{1}{\Delta^2} \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right) \\ &\geqslant (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta}, \end{aligned} \tag{5.2}$$

where the first inequality follows from Lemma 4.4 and the fact that $\varepsilon = (1/\Delta)^{\Theta(1)}$ is sufficiently small, and the last step uses the condition that $|N(u) \cap I_t| \leqslant 2d\sqrt{\gamma} = o(\Delta)$. Hence we have

$$\begin{aligned} \mathbf{E}\left[\, |I_{t+1} \setminus I_t| \,\right] \geqslant \mathbf{E}\left[\, |I_{t+1} \cap B| \,\right] &= \sum_{u \in B} p_{t+1}(u) \geqslant \sum_{u \in B} (1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \\ &= (1 - o(1)) \cdot \frac{e(B, I_t)}{\Delta} \geqslant (1 - o(1)) \cdot |I_t|. \end{aligned}$$

Since $|I_{t+1} \setminus I_t| \leqslant |I_t|$, it follows by using Markov's inequality (applied to $|I_t| - |I_{t+1} \setminus I_t|$) that $\mathbf{Pr}\left[\, |I_{t+1}| \geqslant (2 - f(n))|I_t| \,\right] \geqslant 1 - g(n)$, where $f(n)$ and $g(n)$ are both functions that tend to zero. Hence the time to reach $|I_t| \geqslant n/\log n$ can be upper bounded by the sum of $\log_{2-f(n)} n$ independent, identically distributed geometric random variables with expectation at most $1 - o(1)$ each. Using the Chernoff bound from Lemma A.1 yields for $\tau \triangleq \log_2 n + o(\log n)$ that $\mathbf{Pr}\left[\, |I_{t+\tau}| > n/\log n \,\right] = 1 - o(1)$.

Phase II $|I_t| \in [n/\log n, n - n/\log n]$. We further divide this phase into the two cases $|I_t| \in [n/\log n, n/2]$ and $|I_t| \in [n/2, n - n/\log n]$. We start with the first case $|I_t| \in [n/\log n, n/2]$.

For any $u \in N(I_t) \setminus I_t$, the probability $p_{t+1}(u)$ that u gets informed in round $t+1$ is lowered bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right)$$

by the same argument as in (5.2). This is then lower bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{2\Delta}$$

since we have $|N(u) \cap I_t| \leqslant \Delta$.

By Lemma 5.3, we have

$$\begin{aligned} e(I_t, U_t) &\geqslant \frac{\mathrm{vol}(I_t) \cdot \mathrm{vol}(U_t)}{\mathrm{vol}(G)} - \lambda \frac{\mathrm{vol}(I_t) \cdot \mathrm{vol}(U_t)}{\mathrm{vol}(G)} \\ &= (1 - o(1)) \cdot \frac{\delta}{2} |I_t|. \end{aligned}$$

Similar to the analysis of Phase I, we can lower bound the expected number of nodes that become informed in round $t + 1$:

$$\begin{aligned} \mathbf{E}\left[\, |I_{t+1} \setminus I_t| \,\right] &\geqslant \sum_{u \in N(I_t) \setminus I_t} p_{t+1}(u) \geqslant (1 - o(1)) \sum_{u \in N(I_t) \setminus I_t} \frac{|N(u) \cap I_t|}{2\Delta} \\ &= (1 - o(1)) \frac{e(I_t, U_t)}{2\Delta} \geqslant \frac{\delta}{8\Delta} |I_t|. \end{aligned}$$

Since $|I_{t+1}| \leqslant 2|I_t|$, we obtain as long as $|I_t| \leqslant n/2$ that there are constants $\alpha, \beta > 0$ so that $\mathbf{Pr}\left[\, |I_{t+1}| \geqslant (1 + \alpha)|I_t| \,\right] \geqslant \beta$. Hence the time to reach $|I_t| \geqslant n/2$ can be upper bounded by the sum of $\log_{1+\alpha}(\log n)$ independent, identically distributed geometric random variables with expectation at most $1/\beta$ each. Using a Chernoff bound for the sum of geometric random variables (see Lemma A.1) yields that with probability $1 - o(1)$, we reach $|I_t| \geqslant n/2$ within at most $o(\log n)$ additional rounds.

Consider now the case $|I_t| \in [n/2, n - n/\log n]$. To analyze this case, we examine the shrinking of $U_t = V \setminus I_t$. Note that for any $u \in U_t$, the probability $p_{t+1}(u)$ that u gets informed in round $t + 1$ is lowered bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{\Delta} \left(1 - \frac{(1 + o(1)) \cdot |N(u) \cap I_t|}{2\Delta}\right)$$

by the same argument as in (5.2). This is then lower bounded by

$$(1 - o(1)) \cdot \frac{|N(u) \cap I_t|}{2\Delta}$$

since we have $|N(u) \cap I_t| \leqslant \Delta$.

Again, as $|U_t| \leqslant n/2$, by Lemma 5.3 we have

$$e(I_t, U_t) \geqslant (1 - o(1)) \cdot \frac{\delta}{2} |U_t|.$$

Let us now compute the expected number of uninformed nodes after one additional round:

$$\begin{aligned} \mathbf{E}\left[\, |U_{t+1}| \,\right] &= \sum_{u \in U_t} (1 - p_{t+1}(u)) \leqslant |U_t| - (1 - o(1)) \sum_{u \in U_t} \left(\frac{|N(u) \cap I_t|}{2\Delta}\right) \\ &= |U_t| - (1 - o(1)) \frac{e(I_t, U_t)}{2\Delta} \leqslant \left(1 - \frac{\delta}{8\Delta}\right) |U_t|. \end{aligned}$$

A simple inductive argument yields for any integer τ that,

$$\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant \left(1-\frac{\delta}{8\Delta}\right)^{\tau}|U_t|,$$

so for $\tau \triangleq \log\log n/\log(1/(1-\frac{\delta}{8\Delta}))+\omega(1)$, where $\omega(1)$ is an arbitrarily slow growing function, we have $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] = o(n/\log n)$. Hence by Markov's inequality, $\mathbf{Pr}\left[\,|U_{t+\tau}| \geqslant n/\log n\,\right] = o(1)$.

Phase III $|I_t| \in [n-n/\log n, n]$. Again, we analyze the shrinking of the set U_t. By Lemma 2.13, for at least $(1-\varepsilon/2)$-fraction of the choices of x_t, it holds that the size of $\{\Gamma(v,x_t) : v \in N(u)\cap I_t\}$ is at least $(1-\varepsilon/2)|N(u)\cap I_t|$. From now on fix x_t such that this event occurs.

For any $u \in U_t$, we have

$$\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] = \mathbf{Pr}\left[\bigwedge_{v\in N(u)\cap I_t}(X_{v\to u}=0)\right].$$

Let F be a subset of $N(u)\cap I_t$ of size $(1-\varepsilon/2)|N(u)\cap I_t|$ such that the map $\Gamma(\cdot,x_t)$ is injective when restricted to F. Note that the function $y \mapsto \left(\mathcal{G}_{\Gamma(v,x_t)}(y)\right)_{v\in F} \in [m]^{|F|}$ is a PRG ε'-fooling $\mathcal{M}_{[m]^{|F|},|F|}$. By Lemma 2.5, the function $y \mapsto \left(\mathcal{G}_{\Gamma(v,x_t)}(y) \bmod \deg(v)\right)_{v\in F}$ is a PRG $(\varepsilon'+|F|\Delta/m)$-fooling $\mathcal{M}_{S,|F|}$ where $S=\prod_{v\in F}[\deg(v)]$.

Then we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \mathbf{Pr}\left[\bigwedge_{v\in F}(X_{v\to u}=0)\right] \leqslant \prod_{v\in F}\mathbf{Pr}\left[\,X_{v\to u}=0\,\right]+\varepsilon'+|F|\Delta/m\\
&\leqslant \prod_{v\in F}\left(1-\frac{1}{\deg(v)}+\varepsilon\right)+\varepsilon'+\Delta^2/m\\
&\leqslant \left(1-\frac{1}{\Delta}+\varepsilon\right)^{(1-\varepsilon/2)|N(u)\cap I_t|}+\varepsilon'+\Delta^2/m,
\end{aligned}$$

where the second inequality follows from the properties of PRGs for combinatorial rectangles, and the third inequality follows from using pairwise independent generators. Since $\varepsilon \leqslant \frac{1}{\Delta}$, a simple induction shows that

$$\left(1-\frac{1}{\Delta}+\varepsilon\right)^{k} \leqslant \left(1-\frac{1}{\Delta}\right)^{k}+k\varepsilon$$

for any $k \geqslant 0$. So we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|}+(1-\varepsilon/2)\cdot|N(u)\cap I_t|\cdot\varepsilon+\varepsilon'+\Delta^2/m\\
&\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u)\cap I_t|}+(1-\varepsilon/2)\cdot\Delta\cdot\varepsilon+\varepsilon'+\Delta^2/m.
\end{aligned}$$

The bound above applies for any choice of x_t such that the size of $\{\Gamma(v,x_t) : v \in N(u)\cap I_t\}$ is at least $(1-\varepsilon/2)|N(u)\cap I_t|$. And the probability of choosing such x_t is at least $1-\varepsilon/2$. So for random x_t, we have

$$\begin{aligned}
\mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] &\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)\cdot|N(u)\cap I_t|}+(1-\varepsilon/2)\cdot\Delta\cdot\varepsilon+\varepsilon'+\Delta^2/m+\varepsilon/2\\
&\leqslant \left(1-\frac{1}{\Delta}\right)^{(1-\varepsilon/2)\cdot|N(u)\cap I_t|}+o(1),
\end{aligned}$$

where we use the fact that $\varepsilon = (1/\Delta)^{\Theta}$ is sufficiently small, and $m = \Theta((\log n)/\varepsilon)$.

By (5.1) it holds that $e(I_t, U_t) \geqslant (1 - \frac{1}{\log n} - o(1)) \cdot \Delta |U_t|$. Let $A \subseteq U_t$ be the set of nodes v for which $|N(v) \cap I_t| \leqslant (1 - \sqrt{\gamma}/2) \cdot \Delta$, where $\gamma \triangleq \frac{1}{\log n} + o(1)$. We assume for a contradiction that $|A| > 2\sqrt{\gamma} \cdot |U_t|$. Hence,

$$\begin{aligned} e(I_t, U_t) &= \sum_{v \in A} |N(v) \cap I_t| + \sum_{v \in U_t \setminus A} |N(v) \cap I_t| \leqslant |A| \cdot (1 - \sqrt{\gamma}/2)\Delta + |U_t \setminus A|\Delta \\ &= |U_t|\Delta - |A|\sqrt{\gamma}\Delta/2 < \left(1 - \frac{1}{\log n} - o(1)\right) \cdot \Delta |U_t|, \end{aligned}$$

which yields the desired contradiction. Hence $|A| \leqslant 2\sqrt{\gamma}|U_t|$. Now define $B \triangleq U_t \setminus A$ so that for each $u \in B$, $|N(v) \cap I_t| > (1 - \sqrt{\gamma}/2)\Delta$ and $|B| \geqslant (1 - 2\sqrt{\gamma})|U_t|$. Using linearity of expectations,

$$\begin{aligned} \mathbf{E}\left[\,|U_{t+1}|\,\right] &\leqslant \sum_{u \in B} \mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] + \sum_{u \in A} \mathbf{Pr}\left[\,u \notin I_{t+1}\,\right] \\ &\leqslant \sum_{u \in B} \left(\left(1 - \frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u) \cap I_t|} + o(1)\right) + \sum_{u \in A} 1 \\ &\leqslant \sum_{u \in B} \left(1 - \frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u) \cap I_t|} + o(|U_t|) + |A| \\ &= \sum_{u \in B} \left(1 - \frac{1}{\Delta}\right)^{(1-\varepsilon/2)|N(u) \cap I_t|} + o(|U_t|). \end{aligned}$$

Using the inequalities that $(1 - 1/k) \leqslant \mathrm{e}^{-1/k}$ for $k \geqslant 1$, $\mathrm{e}^x \leqslant 1 + 2x$ for sufficiently small constant $x > 0$, and the condition that $|N(u) \cap I_t| \geqslant (1 - \sqrt{\gamma}/2) \cdot \Delta$ for $u \in B$, we get

$$\begin{aligned} \mathbf{E}\left[\,|U_{t+1}|\,\right] &\leqslant \sum_{u \in B} \mathrm{e}^{-(1-\varepsilon/2)|N(u) \cap I_t|/\Delta} + o(|U_t|) \leqslant \sum_{u \in B} \mathrm{e}^{-(1 - \sqrt{\gamma}/2 - o(1))} + o(|U_t|) \\ &= \sum_{u \in B} \mathrm{e}^{-1} \cdot \mathrm{e}^{\sqrt{\gamma}/2 + o(1)} + o(|U_t|) \leqslant \sum_{u \in B} \mathrm{e}^{-1} \cdot (1 + \sqrt{\gamma} + o(1)) + o(|U_t|) \\ &= (1 + o(1)) \cdot \mathrm{e}^{-1} \cdot |U_t|. \end{aligned}$$

By induction, it follows that for any step $\tau > 0$, $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant ((1 + o(1)) \cdot \mathrm{e}^{-1})^{\tau} \cdot |U_t|$. We choose $\tau \triangleq -\log_{(1+o(1)) \cdot \mathrm{e}^{-1}}(n) = \ln n + o(\log n)$ and obtain that $\mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant (1/\log n)$. So $\mathbf{Pr}\left[\,|U_{t+\tau}| \geqslant 1\,\right] \leqslant \mathbf{E}\left[\,|U_{t+\tau}|\,\right] \leqslant 1/\log n$. ■

Acknowledgement. We are grateful to Chris Umans for many hours of stimulating discussion. We would like to thank Avi Wigderson for helpful discussion at the early stage of our paper.

References

[1] M. Ajtai, J. Komlos, and E. Szemeredi. Deterministic simulation in logspace. In *19th Annual ACM Symposium on Theory of Computing (STOC'87)*, pages 132–140, 1987.

[2] S. Angelopoulos, B. Doerr, A. Huber, and K. Panagiotou. Tight bounds for quasirandom rumor spreading. *Electr. J. Comb.*, 16(1), 2009.

[3] R. Armoni, M. Saks, A. Wigderson, and S. Zhou. Discrepancy sets and pseudorandom generators for combinatorial rectangles. In *37th Annual IEEE Symposium on Foundations of Computer Science (FOCS'96)*, pages 412–421, 1996.

[4] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In *16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*, pages 301–310, 2005.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking*, 52(6): 2508–2530, 2006.

[6] J. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143 – 154, 1979.

[7] K. Censor-Hillel and H. Shachnai. Fast information spreading in graphs with large weak conductance. In *43rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*, pages 440–448, 2011.

[8] K. Censor-Hillel, B. Haeupler, J. A. Kelner, and P. Maymounkov. Global computation in a poorly connected world: fast rumor spreading with no dependence on conductance. In *44th Annual ACM Symposium on Theory of Computing (STOC'12)*, pages 961–970, 2012.

[9] F. Chierichetti, S. Lattanzi, and A. Panconesi. Almost tight bounds on rumour spreading by conductance. In *42nd Annual ACM Symposium on Theory of Computing (STOC'10)*, pages 399–408, 2010.

[10] F. R. K. Chung. Spectral graph theory. *Regional Conference Series in Mathematics, American Mathematical Society*, 92:1–212, 1997.

[11] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *6th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'87)*, pages 1–12, 1987.

[12] B. Doerr and M. Fouz. A time-randomness tradeoff for quasi-random rumour spreading. *Electronic Notes in Discrete Mathematics*, 34:335–339, 2009.

[13] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*, pages 773–781, 2008.

[14] B. Doerr, T. Friedrich, and T. Sauerwald. Quasirandom rumor spreading: Expanders, push vs. pull and robustness. In *36th International Colloquium on Automata, Languages, and Programming (ICALP'09)*, pages 366–377, 2009.

[15] R. Elsässer and T. Sauerwald. Broadcasting vs. mixing and information dissemination on cayley graphs. In *24th International Symposium on Theoretical Aspects of Computer Science (STACS'07)*, pages 163–174. 2007.

[16] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[17] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Veličkovic. Approximations of general independent distributions. In *24th Annual ACM Symposium on Theory of Computing (STOC'92)*, pages 10–16, 1992.

[18] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.

[19] N. Fountoulakis and A. Huber. Quasirandom rumor spreading on the complete graph is as fast as randomized rumor spreading. *SIAM Journal on Discrete Mathematics*, 23(4): 1964–1991, 2009.

[20] N. Fountoulakis, A. Huber, and K. Panagiotou. Reliable broadcasting in random networks and the effect of density. In *29th IEEE Conference on Computer Communications (INFOCOM'10)*, pages 2552–2560, 2010.

[21] G. Giakkoupis. Tight bounds for rumor spreading in graphs of a given conductance. In

28th International Symposium on Theoretical Aspects of Computer Science (STACS'11), pages 57–68, 2011.

[22] G. Giakkoupis and T. Sauerwald. Rumor spreading and vertex expansion. In *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*, pages 1623–1641, 2012.

[23] G. Giakkoupis and P. Woelfel. On the randomness requirements of rumor spreading. In *22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*, pages 449–461, 2011.

[24] G. Giakkoupis, T. Sauerwald, H. Sun, and P. Woelfel. Low randomness rumor spreading via hashing. In *29th International Symposium on Theoretical Aspects of Computer Science (STACS'12)*, pages 314–325, 2012.

[25] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*, 2012.

[26] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Joural of ACM*, 56(4):20:1–20:34, July 2009.

[27] B. Haeupler. Simple, fast and deterministic gossip and rumor spreading. In *24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13)*, pages 705–716, 2013.

[28] M. Harchol-Balter, F. T. Leighton, and D. Lewin. Resource discovery in distributed networks. In *18th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'99)*, pages 229–237, 1999.

[29] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *26th Annual ACM Symposium on Theory of Computing (STOC'94)*, pages 356–364.

[30] C.-J. Lu. Improved pseudorandom generators for combinatorial rectangles. *Combinatorica*, 22(3):417–434, 2002.

[31] D. Mosk-Aoyama and D. Shah. Fast distributed algorithms for computing separable functions. *IEEE Transactions on Information Theory*, 54(7):2997–3007, 2008.

[32] N. Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12 (4):449–461, 1992.

[33] N. Nisan and D. Zuckerman. Randomness is linear in space. *Joural of Computer and System Sciences*, 52(1):43–52, 1996.

[34] A. Ta-Shma, C. Umans, and D. Zuckerman. Lossless condensers, unbalanced expanders, and extractors. *Combinatorica*, 27(2):213–240, 2007.

[35] R. van Renesse, Y. Minsky, and M. Hayden. A gossip-style failure detection service. In *15th IFIP Intl. Conf. on Distributed Systems Platforms (Middleware)*, pages 55–70, 1998.

A Useful Lemmas

Lemma A.1. *Fix any $0 < p < 1$ and let $X_1, \ldots, X_n$ be independent geometric random variables on $\mathbb{N}$ with $\mathbf{Pr}[X_i = k] = (1-p)^{k-1}p$ for every $k \in \mathbb{N}$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = \mathbf{E}[X]$. Then it holds for all $\beta > 0$ that*

$$\mathbf{Pr}[X \geqslant (1+\beta)\mu] \leqslant \mathrm{e}^{-n\beta^2/(2(1+\beta))}.$$

B Existential Proof

In this section we show that $O(\log n)$ random bits are sufficient in rumor spreading for many classes of graphs (e.g. complete graphs, strong expanders, graphs with good conductance, etc.)

if we do not care about the computational complexity. We will prove the following general statement:

Lemma B.1. *Let $\mathcal{C}$ be a class of graphs on n nodes with no multi-edges. Let $T' = n^{O(1)}$ be an upper bound of spreading time. Suppose the spreading time for any graph in $\mathcal{C}$ is at most T with probability p for fully-random push protocol. Then there exists a (non-explicit) function*

$$f : \{0,1\}^\ell \times [n] \times [T'] \times [\Delta] \to [\Delta]$$

such that

1. *$f(x,u,t,d) \in [d]$ for all $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T'] \times [\Delta]$.*
2. *$\ell = \max\{\log\log|C|, \log n + \log\Delta + \log\log\Delta\} + 2\log(1/\varepsilon) + O(1)$.*
3. *for x uniformly chosen from $\{0,1\}^\ell$, the spreading time for any graph $G \in \mathcal{C}$ is at most T with probability $p-\varepsilon$ if node u uses $f(x,u,t,\deg(u)) \in [\deg(u)]$ as the index of its receiver in its adjacency list in round t.*

In particular, ℓ is bounded by $2\log n + \log\log n + 2\log(1/\varepsilon) + O(1)$ since $|\mathcal{C}| \leqslant 2^{n^2}$ and $\Delta \leqslant n$.

Proof. Choose $f(x,u,t,d) \in [d]$ independently and uniformly at random for each $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T'] \times [\Delta]$. Fix a graph $G \in \mathcal{C}$, an initial node in $[n]$. For each node u in the graph of degree $\deg(u)$, there are $\deg(u)!$ possible orders of neighbors of u in its adjacency list. We also fix the order for each node u. Observe that for any fixed x, the random variables $f(x,u,t,\deg(u))$ for all pairs (u,t) are independent and uniformly distributed. Let $I(x)$ be the indicator random variable that equals 1 if the spreading time of G is at most T when node u uses $f(x,u,t,\deg(u))$ to decide its receiver in round t. Then $\mathbf{Pr}_f[\,I(x)=1\,] \geqslant p$ for any x and hence $\mathbf{E}_f[\,I(x)\,] \geqslant p$. Also note that $I(x)$'s are independent. By the Chernoff bound it holds that

$$\mathbf{Pr}_f\left[\left|2^{-\ell}\sum_x I(x) - 2^{-\ell}\sum_x \mathbf{E}_f[\,I(x)\,]\right| \geqslant \varepsilon\right] \leqslant 2\exp(-2^\ell\varepsilon^2/4).$$

So with probability at least $1-2\exp(-2^\ell\varepsilon^2/4)$, we have $\mathbf{E}_x[\,I(x)\,] \geqslant \mathbf{E}_x[\,\mathbf{E}_f[\,I(x)\,]\,] - \varepsilon \geqslant p-\varepsilon$. By the union bound, the probability that $\mathbf{E}_x[\,I(x)\,] \geqslant p-\varepsilon$ holds for all graphs in $\mathcal{C}$, arbitrary neighboring list of nodes, and all start nodes is at least

$$1 - n|\mathcal{C}| \cdot (\Delta!)^n \cdot 2\exp(-2^\ell\varepsilon^2/4)$$

which is greater than zero for sufficiently large $\ell = \max\{\log\log|C|, \log n + \log\Delta + \log\log\Delta\} + 2\log(1/\varepsilon) + O(1)$. So there exists one function f such that $\mathbf{E}_x[\,I(x)\,] \geqslant p-\varepsilon$ holds for all graphs in $\mathcal{C}$, i.e. the spreading time for any graph $G \in \mathcal{C}$ is at most T with probability $p-\varepsilon$ over the choices of x, if node u uses $f(x,u,t,\deg(u)) \in [\deg(u)]$ to choose its receiver in round t. ■

The same result also hold for pull protocols and push-pull protocols, and can be shown using similar arguments.

The following result follows from Lemma B.1 directly.

Corollary B.2 (Existential Result)**.** *Let $\mathcal{G} = \{G_n\}_{n\geqslant 1}$ be a family of graphs such that for any $G_n \in \mathcal{G}$ with n nodes the truly random protocol finishes in $T = n^{O(1)}$ rounds with high probability. Then there is a protocol which finishes in T rounds with high probability and uses $3\log n$ random bits in total.*

C Lower Bounds on Randomness Complexity

We address the randomness requirement of rumor spreading protocols. We first introduce the pull model, which is a symmetric version of the push model, and the formal description is as follows: In round $t \geqslant 0$, every node u that does not yet have the rumor selects a neighbor v uniformly at random and asks for the rumor, and gets the rumor if v received the rumor before. In the push-pull model, in every round t, every node u chooses a random neighbor to perform *push* if node u has the rumor, or perform *pull* if u has not received the rumor.

We prove the following lower bound on the number of random bits needed for any protocol in the push-pull model:

Theorem C.1. *Let G be any graph with n nodes and sufficiently large minimum degree $\delta = \Omega(\log n)$. Then any protocol in the push-pull model that is oblivious of the order of adjacency lists of G and informs at least half of the nodes of G in T rounds with nonzero probability has to use more than $\log\delta - \log T - 2$ random bits. In particular, $\Theta(\log n)$ random bits are necessary when $\delta = \Theta(n)$ and $T = O(n^{1-\varepsilon})$ for some constant $\varepsilon > 0$.*

Here we even allow the protocol access to the ID of the initial node and the structure of G, i.e., the sets of neighbors of nodes as *unordered sets*. All we assume is that the protocol is oblivious of the the *order* of the adjacency lists.

Proof. Suppose $V[G] = [n]$. Let Δ be the maximum degree of G and s be the initial node. We first claim that there exists a subset of nodes S of size $n/2$ (for simplicity assume n is even) such that $\deg(u)/4 \leqslant |S \cap N(u)| \leqslant 3\deg(u)/4$ for all $u \in [n]$: If we pick a random subset S of size $n/2$, then for any fixed u the condition $\deg(u)/4 \leqslant |S \cap N(u)| \leqslant 3\deg(u)/4$ holds, by the Chernoff bound, with probability at least $1 - \mathrm{e}^{-\Theta(\delta)} > 1 - 1/n$ for $\delta = \Omega(\log n)$ sufficiently large. The claim then follows by taking the union bound. Pick such a subset S with the claimed property. Note that $[n] \setminus S$ has the same property. We may therefore assume $s \in S$ by swapping S and $[n] \setminus S$ if necessary.

A protocol for G using ℓ random bits in T rounds is uniquely characterized by a pair of functions

$$f_1, f_2 : \{0,1\}^\ell \times [n] \times [T] \times [\Delta] \to [\Delta]$$

satisfying $f_1(x,u,t,d), f_2(x,u,t,d) \in [d]$ for all $(x,u,t,d) \in \{0,1\}^\ell \times [n] \times [T] \times [\Delta]$, in the sense that given the random string x, node u chooses a neighbor with index $f_1(x,u,t,\deg(u))$ (resp. $f_2(x,u,t,\deg(u))$) in its adjacency list to push (resp. pull) the message in round t if it is informed (resp. uninformed). For each $u \in [n]$, define $I_u \subseteq [n]$ as

$$I_u = \begin{cases} \{f_1(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\} & u \in S \\ \{f_2(x,u,t,\deg(u)) : x \in \{0,1\}^\ell, t \in [T]\} & u \notin S. \end{cases}$$

Assume to the contrary that $\ell \leqslant \log\delta - \log T - 2$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta/4 \leqslant \min\{|S \cap N(u)|, |([n] \setminus S) \cap N(u)|\}$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $S \cap N(u)$ if $u \in S$, or in $([n] \setminus S) \cap N(u)$ if $u \in [n] \setminus S$. Then in the rumor spreading process, nodes in S push messages only to those also in S, and nodes in $[n] \setminus S$ pull messages only from those also in $[n] \setminus S$. As $s \in S$, the nodes in $[n] \setminus S$ never get informed. ∎

For the push model and the pull model we may drop the assumption that $\delta = \Omega(\log n)$ is sufficiently large, and also simplify the proof.

Theorem C.2. *Let G be any graph with n nodes. Then any protocol in the push model that is oblivious of the order of adjacency lists of G and informs all the nodes of G in T rounds with nonzero probability has to use more than $\log(\delta - 1) - \log T$ random bits.*

Proof. The protocol is now characterized by a single function f_1 describing how rumors are pushed. Define $I_u = \{f_1(x, u, t, \deg(u)) : x \in \{0,1\}^\ell, t \in [T]\}$ for each $u \in [n]$. Pick $v \in [n] \setminus \{s\}$. Assume to the contrary that $\ell \leqslant \log(\delta - 1) - \log T$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta - 1 \leqslant |N(u) \setminus \{v\}|$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $N(u) \setminus \{v\}$. Then the node v never gets informed. ■

Theorem C.3. *Let G be any graph with n nodes. Then any protocol in the pull model that is oblivious of the order of adjacency lists of G and informs more than one node of G in T rounds with nonzero probability has to use more than $\log(\delta - 1) - \log T$ random bits.*

Proof. The protocol is now characterized by a single function f_2 describing how rumors are pulled. Define $I_u = \{f_2(x, u, t, \deg(u)) : x \in \{0,1\}^\ell, t \in [T]\}$ for each $u \in [n]$. Assume to the contrary that $\ell \leqslant \log(\delta - 1) - \log T$. Then the size of I_u is at most $2^\ell \cdot T \leqslant \delta - 1 \leqslant |N(u) \setminus \{s\}|$ for each $u \in [n]$. So it is possible to order the adjacency list of each $u \in [n]$ such that the neighbors picked by u using index set I_u are all in $N(u) \setminus \{s\}$. Then the nodes in $[n] \setminus \{s\}$ never get informed. ■

D Omitted Proofs from Section 3

Lemma 3.7 (from page 9)**.** $\|\mathbf{M}\|_2^2 \leqslant \|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty$.

Proof. Choose $\mathbf{u} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ such that $\|\mathbf{uM}\|_2 = \|\mathbf{u}\|_2 \|\mathbf{M}\|_2$. Note that

$$\|\mathbf{uM}\|_2^2 = \langle \mathbf{uM}, \mathbf{uM} \rangle = \langle \mathbf{u}, \mathbf{uMM}^\intercal \rangle \leqslant \lambda_1 \langle \mathbf{u}, \mathbf{u} \rangle = \lambda_1 \|\mathbf{u}\|_2^2$$

where λ_1 is the largest (absolute) eigenvalue of the symmetric matrix $\mathbf{MM}^\intercal$. Let $\mathbf{x}$ be the eigenvector of $\mathbf{MM}^\intercal$ associated with λ_1. We have

$$\lambda_1 \|\mathbf{x}\|_1 = \|\lambda_1 \mathbf{x}\|_1 = \|\mathbf{xMM}^\intercal\|_1 \leqslant \|\mathbf{xM}\|_1 \|\mathbf{M}^\intercal\|_1 \leqslant \|\mathbf{x}\|_1 \|\mathbf{M}\|_1 \|\mathbf{M}^\intercal\|_1 = \|\mathbf{x}\|_1 \|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty.$$

So $\|\mathbf{M}\|_2^2 = \|\mathbf{uM}\|_2^2 / \|\mathbf{u}\|_2^2 \leqslant \lambda_1 \leqslant \|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty$. ■

Lemma 3.16 (from page 15)**.** *For any vector $\mathbf{u} \in \mathbb{R}^{n^2}$, let $\mathbf{u} = \mathbf{u}^\parallel + \mathbf{u}^\perp$ be the (non-orthogonal) decomposition of $\mathbf{u}$ such that $\mathbf{u}^\parallel \parallel \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and $\mathbf{u}^\perp \perp \mathbf{1}$. Then the following statements hold:*

(a) Write $\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2} = \mathbf{u}_1 + \mathbf{u}_2$ such that $\mathbf{u}_1 \parallel \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$ and $\mathbf{u}_2 \perp \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$. Then $\mathbf{u}^\parallel = \mathbf{u}_1(\mathbf{D} \otimes \mathbf{D})^{1/2}$ and $\mathbf{u}^\perp = \mathbf{u}_2(\mathbf{D} \otimes \mathbf{D})^{1/2}$.

(b) $\|\mathbf{u}^\parallel\|_2 \leqslant \beta \|\mathbf{u}\|_2$ and $\|\mathbf{u}^\perp\|_2 \leqslant \beta \|\mathbf{u}\|_2$.

(c) If $\mathbf{u} \in \mathbb{R}^{n^2}$ represents a probability distribution (i.e., all entries of $\mathbf{u}$ are non-negative and $\|\mathbf{u}\|_1 = 1$), then $\mathbf{u}^\parallel = \boldsymbol{\pi} \otimes \boldsymbol{\pi}$.

Proof. For Statement (a), it is easy to verify that $\mathbf{u}_1(\mathbf{D} \otimes \mathbf{D})^{1/2} \parallel \boldsymbol{\pi} \otimes \boldsymbol{\pi}$ and $\mathbf{u}_2(\mathbf{D} \otimes \mathbf{D})^{1/2} \perp \mathbf{1}$. The orthogonal decomposition of $\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}$ into $\mathbf{u}_1$ and $\mathbf{u}_2$ is unique, which implies the uniqueness in (a).

Statement (b) follows from (a) as

$$\|\mathbf{u}^\parallel\|_2 \leqslant \|\mathbf{u}_1\|_2 \|(\mathbf{D} \otimes \mathbf{D})^{1/2}\|_2 \leqslant \|\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}\|_2 \cdot \Delta \leqslant \beta \|\mathbf{u}\|_2$$

and similarly

$$\|\mathbf{u}^\perp\|_2 \leqslant \|\mathbf{u}_2\|_2 \|(\mathbf{D} \otimes \mathbf{D})^{1/2}\|_2 \leqslant \|\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}\|_2 \cdot \Delta \leqslant \beta \|\mathbf{u}\|_2.$$

Statement (c) follows from the facts that

$$\langle \mathbf{u}^\parallel, \mathbf{1} \rangle = \langle \mathbf{u}_1, \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2} \rangle = \langle \mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}, \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2} \rangle = 1 = \langle \boldsymbol{\pi} \otimes \boldsymbol{\pi}, \mathbf{1} \rangle$$

and $\mathbf{u}^\parallel \parallel \boldsymbol{\pi} \otimes \boldsymbol{\pi}$. ■

Lemma 3.17 (from page 15). *For any vector* $\mathbf{u} \in \mathbb{R}^{n^2}$ *parallel to* $\boldsymbol{\pi} \otimes \boldsymbol{\pi}$*, it holds that* $\|(\mathbf{uQ})^{\perp}\|_2 \leqslant \beta^4 n^{-1/2}\|\mathbf{u}\|_2$.

Proof. Without loss of generality assume $\mathbf{u} = \mathbf{1}(\mathbf{D} \otimes \mathbf{D})$. Let

$$\mathbf{E} \triangleq \mathbf{P} - \mathbf{M}_G \otimes \mathbf{M}_G.$$

Then $\mathbf{Q} = \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) + \frac{\mathbf{E}}{4}$. By the definition of matrix $\mathbf{E}$, we have

$$\mathbf{E}_{(u,v)(w,x)} = \begin{cases} 0 & u \neq v, \\ (\mathbf{M}_G)_{(u,w)} - (\mathbf{M}_G \otimes \mathbf{M}_G)_{(u,v)(w,x)} & u = v, w = x, \\ -(\mathbf{M}_G \otimes \mathbf{M}_G)_{(u,v)(w,x)} & u = v, w \neq x. \end{cases}$$

Then for any $w, x \in V[G]$, we have

$$|(\mathbf{uE})_{(w,x)}| \leqslant \Delta^2|(\mathbf{1M}_G)_w| + \Delta^2|(\mathbf{1}(\mathbf{M}_G \otimes \mathbf{M}_G))_{(w,x)}| \leqslant \Delta^2\|\mathbf{M}_G\|_\infty + \Delta^2\|\mathbf{M}_G \otimes \mathbf{M}_G\|_\infty \leqslant 2\beta\Delta^2.$$

So $\|\mathbf{uE}\|_\infty \leqslant 2\beta\Delta^2$. We also have

$$\sum_{w \neq x} |(\mathbf{uE})_{(w,x)}| \leqslant \sum_{u,w,x} \Delta^2(\mathbf{M}_G \otimes \mathbf{M}_G)_{(u,u)(w,x)} \leqslant n\Delta^2\|\mathbf{M}_G \otimes \mathbf{M}_G\|_1 = n\Delta^2.$$

So $\|\mathbf{uE}\|_1 = \sum_{w,x} |(\mathbf{uE})_{(w,x)}| \leqslant n \cdot 2\beta\Delta^2 + n\Delta^2$. By Hölder's inequality, we have

$$\|\mathbf{uE}\|_2^2 = \langle \mathbf{uE}, \mathbf{uE} \rangle \leqslant \|\mathbf{uE}\|_1\|\mathbf{uE}\|_\infty \leqslant (4\beta^2 + 2\beta)n\Delta^4.$$

As $\mathbf{Q} = \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) + \frac{\mathbf{E}}{4}$ and $\mathbf{1}(\mathbf{D} \otimes \mathbf{D})$ is fixed by $\left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right) \otimes \left(\frac{\mathbf{I}+\mathbf{M}_G}{2}\right)$, we have $(\mathbf{uQ})^{\perp} = (\mathbf{uE})^{\perp}/4$. Therefore

$$\|(\mathbf{uQ})^{\perp}\|_2 \leqslant (\beta/4)\|\mathbf{uE}\|_2 \leqslant (\beta/4)\sqrt{(4\beta^2 + 2\beta)n\Delta^4} \leqslant \beta^2 n^{1/2}\Delta^2.$$

Finally note that $\|\mathbf{u}\|_2 = \|\mathbf{1}(\mathbf{D} \otimes \mathbf{D})\|_2 \geqslant n\delta^2$ and hence $\|(\mathbf{uQ})^{\perp}\|_2 \leqslant \beta^4 n^{-1/2}\|\mathbf{u}\|_2$. ∎

Lemma 3.18 (from page 15). *For any vector* $\mathbf{u} \in \mathbb{R}^{n^2}$ *orthogonal to* $\mathbf{1}$*, it holds that* $\mathbf{uQ} \perp \mathbf{1}$ *and*

$$\left\|\mathbf{uQD}'^{-1/2}\right\|_2 \leqslant (1 - \alpha/8)\left\|\mathbf{uD}'^{-1/2}\right\|_2.$$

Proof. Since $\mathbf{Q}$ is stochastic, we have $\langle \mathbf{uQ}, \mathbf{1} \rangle = \langle \mathbf{u}, \mathbf{1Q}^{\intercal} \rangle = \langle \mathbf{u}, \mathbf{1} \rangle = 0$. To prove the second claim, we look at each summand in

$$4\mathbf{Q} = \mathbf{I} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{M}_G + \mathbf{M}_G \otimes \mathbf{I} + \mathbf{P}$$

individually. Clearly $\|\mathbf{u}(\mathbf{I} \otimes \mathbf{I})\mathbf{D}'^{-1/2}\|_2 = \|\mathbf{uD}'^{-1/2}\|_2$. Write $\mathbf{I} \otimes \mathbf{M}_G = \mathbf{M} + \mathbf{M}'$ such that

$$\mathbf{M}_{(u,v)(w,x)} = \begin{cases} (\mathbf{I} \otimes \mathbf{M}_G)_{(u,v)(w,x)} & w = x \\ 0 & w \neq x \end{cases} \qquad \mathbf{M}'_{(u,v)(w,x)} = \begin{cases} 0 & w = x \\ (\mathbf{I} \otimes \mathbf{M}_G)_{(u,v)(w,x)} & w \neq x. \end{cases}$$

Note that each row (u, v) of $\mathbf{M}$ contains at most one nonzero entry, namely the one at the column (u, u) whose value is at most δ^{-1}. So $\|\mathbf{M}\|_1 \leqslant \delta^{-1}$. We also have $\|\mathbf{M}\|_\infty \leqslant \|\mathbf{I} \otimes \mathbf{M}_G\|_\infty \leqslant \beta$, and hence $\|\mathbf{M}\|_2 \leqslant \sqrt{\|\mathbf{M}\|_1\|\mathbf{M}\|_\infty} \leqslant \beta^{1/2}\delta^{-1/2}$. Write

$$\mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)\mathbf{D}'^{-1/2} = \mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)(\mathbf{D} \otimes \mathbf{D})^{-1/2} + \mathbf{r}.$$

Here

$$\mathbf{r} = \mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)(\mathbf{D}'^{-1/2} - (\mathbf{D} \otimes \mathbf{D})^{-1/2}) = \mathbf{u}\mathbf{M}(\mathbf{D}'^{-1/2} - (\mathbf{D} \otimes \mathbf{D})^{-1/2}),$$

since the columns (w, x) of $\mathbf{I} \otimes \mathbf{M}_G - \mathbf{M} = \mathbf{M}'$ are zero if $w = x$, whereas the rows (u, v) of $\mathbf{D}'^{-1/2} - (\mathbf{D} \otimes \mathbf{D})^{-1/2}$ are zero if $u \neq v$. We have

$$\begin{aligned}\|\mathbf{r}\|_2 &\leqslant \|\mathbf{u}\|_2 \|\mathbf{M}\|_2 \left\|\mathbf{D}'^{-1/2} - (\mathbf{D} \otimes \mathbf{D})^{-1/2}\right\| \\ &\leqslant \|\mathbf{u}\|_2 \cdot (\beta^{1/2}\delta^{-1/2}) \cdot \delta^{-1} \leqslant \beta^{3/2}\delta^{-1/2} \left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2.\end{aligned}$$

Symmetrically, we have

$$\mathbf{u}(\mathbf{M}_G \otimes \mathbf{I})\mathbf{D}'^{-1/2} = \mathbf{u}(\mathbf{M}_G \otimes \mathbf{I})(\mathbf{D} \otimes \mathbf{D})^{-1/2} + \mathbf{r}'$$

with $\|\mathbf{r}'\|_2 \leqslant \beta^{3/2}\delta^{-1/2}\|\mathbf{u}\mathbf{D}'^{-1/2}\|_2$. Note that

$$\mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)(\mathbf{D} \otimes \mathbf{D})^{-1/2} + \mathbf{u}(\mathbf{M}_G \otimes \mathbf{I})(\mathbf{D} \otimes \mathbf{D})^{-1/2} = \mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}(\mathbf{I} \otimes \mathbf{N}_G + \mathbf{N}_G \otimes \mathbf{I}).$$

Recall that $\mathbf{N}_G$ has n normalized orthogonal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$ associated with n real eigenvalues $1 = \lambda_1 > 1 - \alpha \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant \alpha - 1$ respectively, and $\mathbf{v}_1$ is parallel to $\mathbf{1}\mathbf{D}^{1/2}$. Then $\mathbf{I} \otimes \mathbf{N}_G + \mathbf{N}_G \otimes \mathbf{I}$ has n^2 normalized orthogonal eigenvectors $\mathbf{v}_i \otimes \mathbf{v}_j$ associated with eigenvalues $\lambda_j + \lambda_i$, $i, j = 1, \ldots, n$. The largest eigenvalue of $\mathbf{I} \otimes \mathbf{N}_G + \mathbf{N}_G \otimes \mathbf{I}$ is 2, corresponding to the eigenvector $\mathbf{v}_1 \otimes \mathbf{v}_1 \parallel \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$, and the second largest (absolute) eigenvalue is bounded by $2 - \alpha$. As $\mathbf{u} \perp \mathbf{1}$, we have $\langle \mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}, \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2} \rangle = \langle \mathbf{u}, \mathbf{1} \rangle = 0$, i.e., $\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2} \perp \mathbf{1}(\mathbf{D} \otimes \mathbf{D})^{1/2}$. Therefore

$$\begin{aligned}&\left\|\mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)(\mathbf{D} \otimes \mathbf{D})^{-1/2} + \mathbf{u}(\mathbf{M}_G \otimes \mathbf{I})(\mathbf{D} \otimes \mathbf{D})^{-1/2}\right\|_2 \\ &\leqslant (2 - \alpha)\left\|\mathbf{u}(\mathbf{D} \otimes \mathbf{D})^{-1/2}\right\|_2 \leqslant (2 - \alpha)\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2\end{aligned}$$

and hence

$$\begin{aligned}\left\|\mathbf{u}(\mathbf{I} \otimes \mathbf{M}_G)\mathbf{D}'^{-1/2} + \mathbf{u}(\mathbf{M}_G \otimes \mathbf{I})\mathbf{D}'^{-1/2}\right\|_2 &\leqslant (2 - \alpha)\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2 + \|\mathbf{r}\|_2 + \|\mathbf{r}'\|_2 \\ &\leqslant (2 - \alpha + 2\beta^{3/2}\delta^{-1/2})\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2.\end{aligned}$$

Finally we look at matrix $\mathbf{P}$. By permutating the rows (resp. columns) of $\mathbf{P}$, we assume its first n rows (resp. n columns) are indexed by the diagonal elements $\{(u, u) : u \in V[G]\}$. By definition, we have

$$\mathbf{P} = \begin{pmatrix}\mathbf{M}_G & \mathbf{0} \\ \mathbf{M}_1 & \mathbf{M}_2\end{pmatrix}$$

where $\begin{pmatrix}\mathbf{M}_1 & \mathbf{M}_2\end{pmatrix}$ are the last $n^2 - n$ rows of $\mathbf{M}_G \otimes \mathbf{M}_G$ (we permutate the rows and columns of $\mathbf{M}_G \otimes \mathbf{M}_G$ in the same way as we did for $\mathbf{P}$). We claim that $\|\mathbf{u}\mathbf{P}\mathbf{D}'^{-1/2}\|_2 \leqslant (1 + 2\beta^2\delta^{-1/2})\|\mathbf{u}\mathbf{D}'^{-1/2}\|_2$ for any vector $\mathbf{u} \in \mathbb{R}^{n^2}$ (not necessarily orthogonal to $\mathbf{1}$): Write $\mathbf{u} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{pmatrix}$ where $\mathbf{u}_1 \in \mathbb{R}^n$ and $\mathbf{u}_2 \in \mathbb{R}^{n^2-n}$, consisting of entries indexed by (u, v), $u = v$ and $u \neq v$ respectively. Also write

$$\mathbf{D} \otimes \mathbf{D} = \begin{pmatrix}\mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2\end{pmatrix}, \qquad \mathbf{D}' = \begin{pmatrix}\mathbf{D}'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2\end{pmatrix}$$

where $\mathbf{D}_1, \mathbf{D}'_1 \in \mathbb{R}^{n \times n}$ and $\mathbf{D}_2 \in \mathbb{R}^{(n^2-n)\times(n^2-n)}$. Here $(\mathbf{D}_1)_{uu} = (\mathbf{D} \otimes \mathbf{D})_{(u,u)(u,u)} = (\deg(u))^2$ and $\mathbf{D}'_1 = \delta \cdot \mathbf{D}$. Then

$$\|\mathbf{u}\mathbf{D}'^{-1/2}\|_2^2 = \|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\|_2^2 + \|\mathbf{u}_2\mathbf{D}_2^{-1/2}\|_2^2$$

and

$$\begin{aligned}\left\|\mathbf{u}\mathbf{P}\mathbf{D}'^{-1/2}\right\|_2^2 &= \left\|\begin{pmatrix}\mathbf{u}_1\mathbf{M}_G\mathbf{D}_1'^{-1/2}+\mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2} & \mathbf{u}_2\mathbf{M}_2\mathbf{D}_2'^{-1/2}\end{pmatrix}\right\|_2^2 \\ &= \left\|\mathbf{u}_1\mathbf{M}_G\mathbf{D}_1'^{-1/2}\right\|_2^2+\left\|\mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2}\right\|_2^2+\left\|\mathbf{u}_2\mathbf{M}_2\mathbf{D}_2'^{-1/2}\right\|_2^2 \\ &\quad+2\left\langle\mathbf{u}_1\mathbf{M}_G\mathbf{D}_1'^{-1/2},\mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2}\right\rangle \\ &= \delta^{-1}\left\|\mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2}\right\|_2^2+\left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}\mathbf{D}'^{-1/2}\right\|_2^2 \\ &\quad+2\delta^{-1/2}\left\langle\mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2},\mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2}\right\rangle.\end{aligned} \tag{D.1}$$

We bound the first term by

$$\delta^{-1}\left\|\mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2}\right\|_2^2=\delta^{-1}\left\|\mathbf{u}_1\mathbf{D}^{-1/2}\mathbf{N}_G\right\|_2^2\leqslant\delta^{-1}\left\|\mathbf{u}_1\mathbf{D}^{-1/2}\right\|_2^2=\left\|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\right\|_2^2 \tag{D.2}$$

and the inequality in the middle holds since the largest (absolute) eigenvalue of $\mathbf{N}_G$ is 1.

For the second term, write

$$\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}\mathbf{D}'^{-1/2}=\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}\otimes\mathbf{D})^{-1/2}+\mathbf{f}$$

where $\mathbf{f}=\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}'^{-1/2}-(\mathbf{D}\otimes\mathbf{D})^{-1/2})$. As $\mathbf{P}$ and $\mathbf{M}_G\otimes\mathbf{M}_G$ disagree only on rows (u,v) for $u=v$, we have

$$\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}\otimes\mathbf{D})^{-1/2}=\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}(\mathbf{M}_G\otimes\mathbf{M}_G)(\mathbf{D}\otimes\mathbf{D})^{-1/2}=\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}(\mathbf{D}\otimes\mathbf{D})^{-1/2}(\mathbf{N}_G\otimes\mathbf{N}_G).$$

Therefore

$$\begin{aligned}\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}\otimes\mathbf{D})^{-1/2}\|_2 &= \left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}(\mathbf{D}\otimes\mathbf{D})^{-1/2}(\mathbf{N}_G\otimes\mathbf{N}_G)\right\|_2 \\ &\leqslant\left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}(\mathbf{D}\otimes\mathbf{D})^{-1/2}\right\|_2 \\ &=\left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2.\end{aligned}$$

Then we bound $\|\mathbf{f}\|_2$. Note that

$$\|\mathbf{f}\|_2=\left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}'^{-1/2}-(\mathbf{D}\otimes\mathbf{D})^{-1/2})\right\|_2=\left\|\mathbf{u}_2\mathbf{M}_1(\mathbf{D}_1'^{-1/2}-\mathbf{D}_1^{-1/2})\right\|_2.$$

We have $\|\mathbf{M}_1\|_\infty\leqslant\|\mathbf{M}_G\otimes\mathbf{M}_G\|_\infty\leqslant\beta^2$. To bound $\|\mathbf{M}_1\|_1$, observe that $\|\mathbf{M}_1\|_1$ is by definition the maximum of the ℓ_1-norm of rows of $\mathbf{M}_1$. Also note that $G\times G$ has no multi-edge and therefore a nonzero entry in the (u,v)-th row of $\mathbf{M}_G\otimes\mathbf{M}_G=(\mathbf{D}\otimes\mathbf{D})^{-1}\mathbf{A}_{G\times G}$ (and hence also $\mathbf{M}_1$) has value $1/(\deg(u)\deg(v))$. So

$$\|\mathbf{M}_1\|_1=\max_{\substack{(u,v)\in V[G]\times V[G]\\u\neq v}}\sum_{(w,w)\in N((u,v))}1/(\deg(u)\deg(v))=\max_{\substack{u,v\in V[G]\\u\neq v}}\sum_{w\in N(u)\cap N(v)}1/(\deg(u)\deg(v)).$$

For any $u,v\in V[G]$, we have $|N(u)\cap N(v)|\leqslant\min\{\deg(u),\deg(v)\}$, which implies $\|\mathbf{M}_1\|_1\leqslant\min\{1/\deg(u),1/\deg(v)\}\leqslant\delta^{-1}$. We then have $\|\mathbf{M}_1\|_2^2\leqslant\|\mathbf{M}_1\|_1\|\mathbf{M}_1\|_\infty\leqslant\beta^2\delta^{-1}$. Then

$$\|\mathbf{f}\|_2\leqslant\|\mathbf{u}_2\|_2\|\mathbf{M}_1\|_2\left\|\mathbf{D}_1'^{-1/2}-\mathbf{D}_1^{-1/2}\right\|_2$$

which is bounded by

$$\|\mathbf{u}_2\|_2\|\mathbf{M}_1\|_2\left\|\mathbf{D}_1'^{-1/2}\right\|_2\leqslant\|\mathbf{u}_2\|_2\cdot(\beta\delta^{-1/2})\cdot\delta^{-1}\leqslant\beta^2\delta^{-1/2}\left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2. \tag{D.3}$$

Therefore

$$\begin{aligned}\left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}\mathbf{D}'^{-1/2}\right\|_2^2 &\leqslant \left(\left\|\begin{pmatrix}\mathbf{0} & \mathbf{u}_2\end{pmatrix}\mathbf{P}(\mathbf{D}\otimes\mathbf{D})^{-1/2}\right\|_2 + \|\mathbf{f}\|_2\right)^2 \\ &\leqslant (1+\beta^2\delta^{-1/2})^2\left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2^2. \end{aligned} \tag{D.4}$$

Then we bound the last term

$$\begin{aligned}2\delta^{-1/2}\left\langle \mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2}, \mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2}\right\rangle &\leqslant 2\delta^{-1/2}\left\|\mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2}\right\|_2\left\|\mathbf{u}_2\mathbf{M}_1\mathbf{D}_1'^{-1/2}\right\|_2 \\ &\leqslant 2\delta^{-1/2}\left\|\mathbf{u}_1\mathbf{M}_G\mathbf{D}^{-1/2}\right\|_2\|\mathbf{u}_2\|_2\|\mathbf{M}_1\|_2\left\|\mathbf{D}_1'^{-1/2}\right\|_2 \\ &\leqslant 2\beta^2\delta^{-1/2}\left\|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\right\|_2\left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2 \\ &\leqslant \beta^2\delta^{-1/2}\left(\left\|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\right\|_2^2 + \left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2^2\right)\end{aligned} \tag{D.5}$$

where we use (D.2) and (D.3) in the third step.

Combining (D.1), (D.2), (D.4) and (D.5), we obtain

$$\begin{aligned}\|\mathbf{u}\mathbf{P}\mathbf{D}'^{-1/2}\|_2^2 &\leqslant \left(1+\beta^2\delta^{-1/2}\right)\left\|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\right\|_2^2 + \left((1+\beta^2\delta^{-1/2})^2+\beta^2\delta^{-1/2}\right)\left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2^2 \\ &\leqslant \left(1+2\beta^2\delta^{-1/2}\right)^2\left(\left\|\mathbf{u}_1\mathbf{D}_1'^{-1/2}\right\|_2^2 + \left\|\mathbf{u}_2\mathbf{D}_2^{-1/2}\right\|_2^2\right) \\ &= \left(1+2\beta^2\delta^{-1/2}\right)^2\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2^2\end{aligned}$$

and hence

$$\left\|\mathbf{u}\mathbf{P}\mathbf{D}'^{-1/2}\right\|_2 \leqslant (1+2\beta^2\delta^{-1/2})\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2,$$

as claimed.

Combining all the results above, we have

$$\begin{aligned}&\left\|\mathbf{u}\mathbf{Q}\mathbf{D}'^{-1/2}\right\|_2 \\ &\leqslant \frac{\left\|\mathbf{u}(\mathbf{I}\otimes\mathbf{I})\mathbf{D}'^{-1/2}\right\|_2 + \left\|\mathbf{u}(\mathbf{I}\otimes\mathbf{M}_G)\mathbf{D}'^{-1/2} + \mathbf{u}(\mathbf{M}_G\otimes\mathbf{I})\mathbf{D}'^{-1/2}\right\|_2 + \left\|\mathbf{u}\mathbf{P}\mathbf{D}'^{-1/2}\right\|_2}{4} \\ &\leqslant \frac{(1+(2-\alpha+2\beta^{3/2}\delta^{-1/2})+(1+2\beta^2\delta^{-1/2}))\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2}{4} \\ &\leqslant (1-\alpha/8)\left\|\mathbf{u}\mathbf{D}'^{-1/2}\right\|_2\end{aligned}$$

where we use the assumption $\alpha \geqslant 8\Delta^2\delta^{-2.5}$. ■

Lemma 3.19 (from page 16)**.** *Let* $\mathbf{u}$ *represent a distribution over* $V[G]\times V[G]$*, and suppose matrix* $\mathbf{M}$ *is a convex combination of products of* $\mathbf{I}\otimes\mathbf{I}$*,* $\mathbf{I}\otimes\mathbf{M}_G$*,* $\mathbf{M}_G\otimes\mathbf{I}$ *and* $\mathbf{P}$*. Then for any* $k\in\mathbb{N}$ *it holds that*

$$\left\|\mathcal{T}_1(\mathbf{u}\mathbf{Q}^k\mathbf{M}) - \boldsymbol{\pi}\right\|_2 \leqslant (1-\alpha/2)^k\beta^{1/2},$$

and

$$\left\|\mathcal{T}_2(\mathbf{u}\mathbf{Q}^k\mathbf{M}) - \boldsymbol{\pi}\right\|_2 \leqslant (1-\alpha/2)^k\beta^{1/2}.$$

Proof. We prove the claim for $\mathcal{T}_1(\mathbf{u}\mathbf{Q}^k\mathbf{M})$ and the case for $\mathcal{T}_2$ is symmetric. As one can check, $\mathcal{T}_1(\mathbf{u}\mathbf{Q}^k\mathbf{M}) = \mathbf{u}'((\mathbf{M}_G+\mathbf{I})/2)^k\mathbf{M}' \in \mathbb{R}^n$ where $\mathbf{u}' = \mathcal{T}_1(\mathbf{u}) \in \mathbb{R}^n$ and $\mathbf{M}'$ is a convex combination of powers of $\mathbf{M}_G$. Write $\mathbf{u}' = \mathbf{u}'_1 + \mathbf{u}'_2$ such that $\mathbf{u}'_1 \parallel \boldsymbol{\pi}$ and $\mathbf{u}'_2 \perp \mathbf{1}$. Then $\mathbf{u}'_1 = \boldsymbol{\pi}$ is fixed by $((\mathbf{M}_G + \mathbf{I})/2)^k\mathbf{M}'$. So $\mathcal{T}_1(\mathbf{u}\mathbf{Q}^k\mathbf{M}) - \boldsymbol{\pi} = \mathbf{u}'_2((\mathbf{M}_G + \mathbf{I})/2)^k\mathbf{M}'$.

It remains to show $\|\mathbf{u}'_2((\mathbf{M}_G+\mathbf{I})/2)^k\mathbf{M}'\|_2 \leqslant (1-\alpha/2)^k\beta^{1/2}$. Let $\mathbf{N} = \mathbf{D}^{1/2}\mathbf{M}'\mathbf{D}^{-1/2}$ which is a convex combination of powers of $\mathbf{N}_G$. As $\mathbf{u}'_2\mathbf{D}^{-1/2} \perp \mathbf{1}\mathbf{D}^{1/2}$, the matrix $((\mathbf{N}_G + \mathbf{I})/2)^k\mathbf{N}_G$ reduces the ℓ_2-norm of $\mathbf{u}'_2\mathbf{D}^{-1/2}$ by at least a factor of $(1-\alpha/2)^{-k}$ by the spectral expansion of G. Then

$$\begin{aligned}
\left\|\mathbf{u}'_2((\mathbf{M}_G + \mathbf{I})/2)^k\mathbf{M}'\right\|_2 &= \left\|\mathbf{u}'_2\mathbf{D}^{-1/2}((\mathbf{N}_G + \mathbf{I})/2)^k\mathbf{N}\mathbf{D}^{1/2}\right\|_2 \\
&\leqslant (1-\alpha/2)^k \left\|\mathbf{u}'_2\mathbf{D}^{-1/2}\right\|_2 \left\|\mathbf{D}^{1/2}\right\|_2 \\
&\leqslant (1-\alpha/2)^k \left\|\mathbf{u}'\mathbf{D}^{-1/2}\right\|_2 \left\|\mathbf{D}^{1/2}\right\|_2 \\
&\leqslant (1-\alpha/2)^k\beta^{1/2} \left\|\mathbf{u}'\right\|_2 \\
&\leqslant (1-\alpha/2)^k\beta^{1/2}.
\end{aligned}$$

The third step uses the fact that $\mathbf{u}'_1\mathbf{D}^{-1/2}$ and $\mathbf{u}'_2\mathbf{D}^{-1/2}$ are orthogonal. The last step uses the fact that $\|\mathbf{u}'\|_2 \leqslant \|\mathbf{u}'\|_1 = 1$. ■

Lemma 3.20 (from page 16)**.** *Let* $\mathbf{u}$ *represent a distribution over* $V[G] \times V[G]$*. Then for any* $k \in \mathbb{N}$ *it holds that*

$$\left\|\mathbf{u}\mathbf{Q}^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi}\right\|_2 \leqslant 8\alpha^{-1}\beta^7 n^{-3/2} + (1-\alpha/8)^k\beta^2.$$

Proof. Note that $(\mathbf{u}\mathbf{Q}^k)^{\parallel} = \boldsymbol{\pi}\otimes\boldsymbol{\pi}$ and hence

$$\mathbf{u}\mathbf{Q}^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi} = (\mathbf{u}\mathbf{Q}^k)^{\perp}.$$

So it suffices to show

$$\left\|(\mathbf{u}\mathbf{Q}^k)^{\perp}\mathbf{D}'^{-1/2}\right\|_2 \leqslant 8\alpha^{-1}\beta^6\delta^{-1}n^{-3/2} + (1-\alpha/8)^k\beta\delta^{-1}.$$

Induct on k. The case $k = 0$ is trivial. For $k > 0$, assume the claim holds for $k' < k$. Let $\mathbf{v} = \mathbf{u}\mathbf{Q}^{k-1}$. Then $\mathbf{v}$ represents a distribution since $\mathbf{Q}^{k-1}$ is stochastic. Write $\mathbf{v} = \mathbf{v}^{\parallel} + \mathbf{v}^{\perp}$. Then $\mathbf{v}^{\parallel} = \boldsymbol{\pi}\otimes\boldsymbol{\pi}$, and

$$\left\|\mathbf{v}^{\parallel}\right\|_2 = \frac{\sum_{u\in V[G]}\deg^2(u)}{(\sum_{u\in V[G]}\deg(u))^2} \leqslant \beta^2 n^{-1}.$$

We have

$$(\mathbf{u}\mathbf{Q}^k)^{\perp}\mathbf{D}'^{-1/2} = (\mathbf{v}\mathbf{Q})^{\perp}\mathbf{D}'^{-1/2} = (\mathbf{v}^{\parallel}\mathbf{Q})^{\perp}\mathbf{D}'^{-1/2} + (\mathbf{v}^{\perp}\mathbf{Q})^{\perp}\mathbf{D}'^{-1/2}.$$

By Lemma 3.17,

$$\left\|(\mathbf{v}^{\parallel}\mathbf{Q})^{\perp}\right\|_2 \leqslant \beta^4 n^{-1/2}\left\|\mathbf{v}^{\parallel}\right\|_2 \leqslant \beta^6 n^{-3/2}$$

and hence

$$\left\|(\mathbf{v}^{\parallel}\mathbf{Q})^{\perp}\mathbf{D}'^{-1/2}\right\|_2 \leqslant \beta^6\delta^{-1}n^{-3/2}.$$

By Lemma 3.18,

$$\left\|\mathbf{v}^{\perp}\mathbf{Q}\mathbf{D}'^{-1/2}\right\|_2 \leqslant (1-\alpha/8)\left\|\mathbf{v}^{\perp}\mathbf{D}'^{-1/2}\right\|_2 = (1-\alpha/8)\left\|(\mathbf{u}\mathbf{Q}^{k-1})^{\perp}\mathbf{D}'^{-1/2}\right\|_2,$$

which is bounded by $(1-\alpha/8)\cdot 8\alpha^{-1}\beta^6\delta^{-1}n^{-3/2}+(1-\alpha/8)^k\beta\delta^{-1}$ by the induction hypothesis. Then

$$\begin{aligned}\left\|(\mathbf{u}\mathbf{Q}^k)^{\perp}\mathbf{D}'^{1/2}\right\|_2 &\leqslant \left\|(\mathbf{v}^{\parallel}\mathbf{Q})^{\perp}\mathbf{D}'^{1/2}\right\|_2+\left\|\mathbf{v}^{\perp}\mathbf{Q}\mathbf{D}'^{1/2}\right\|_2\\ &\leqslant \beta^6\delta^{-1}n^{-3/2}+(1-\alpha/8)\cdot 8\alpha^{-1}\beta^6\delta^{-1}n^{-3/2}+(1-\alpha/8)^k\beta\delta^{-1}\\ &\leqslant 8\alpha^{-1}\beta^6\delta^{-1}n^{-3/2}+(1-\alpha/8)^k\beta\delta^{-1}\end{aligned}$$

as desired. ■

Lemma 3.21 (from page 16)**.** *Let* $\mathbf{u}$ *represent a distribution over* $V[G]\times V[G]$*. And suppose matrix* $\mathbf{M}$ *satisfies*

$$\mathbf{M}=\left(\frac{\mathbf{M}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}\quad or\quad \mathbf{M}=\left(\mathbf{I}\otimes\frac{\mathbf{M}_G+\mathbf{I}}{2}\right)^{k'}$$

for some $k'\in\mathbb{N}$*. Then for* $k\in\mathbb{N}$*, it holds that*

$$\begin{aligned}&\|\mathbf{u}\mathbf{Q}^k\mathbf{M}-\boldsymbol{\pi}\otimes\boldsymbol{\pi}\|_2\\ &\leqslant (1-\alpha/2)^k\beta^{3/2}+(1-\alpha/2)^{k'}\cdot 8\alpha^{-1}\beta^8n^{-3/2}+(1-\alpha/2)^{k'}(1-\alpha/8)^k\beta^3.\end{aligned}$$

Proof. We prove the claim for the case $\mathbf{M}=\left(\frac{\mathbf{M}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}$ and the other one is symmetric. Note that

$$(\mathbf{u}\mathbf{Q}^k)^{\parallel}\mathbf{M}=(\boldsymbol{\pi}\otimes\boldsymbol{\pi})\mathbf{M}=\boldsymbol{\pi}\otimes\boldsymbol{\pi}$$

and hence $\mathbf{u}\mathbf{Q}^k-\boldsymbol{\pi}\otimes\boldsymbol{\pi}=(\mathbf{u}\mathbf{Q}^k)^{\perp}\mathbf{M}$. Write $\mathbf{u}\mathbf{Q}^k(\mathbf{D}\otimes\mathbf{D})^{-1/2}=\mathbf{u}_1+\mathbf{u}_2$ where $\mathbf{u}_1\parallel\mathbf{1}(\mathbf{D}\otimes\mathbf{D})^{1/2}$ and $\mathbf{u}_2\perp\mathbf{1}(\mathbf{D}\otimes\mathbf{D})^{1/2}$. Then $(\mathbf{u}\mathbf{Q}^k)^{\perp}=\mathbf{u}_2(\mathbf{D}\otimes\mathbf{D})^{1/2}$ by definition. Therefore

$$(\mathbf{u}\mathbf{Q}^k)^{\perp}\mathbf{M}=\mathbf{u}_2(\mathbf{D}\otimes\mathbf{D})^{1/2}\left(\frac{\mathbf{M}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}=\mathbf{u}_2\left(\frac{\mathbf{N}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}(\mathbf{D}\otimes\mathbf{D})^{1/2}.$$

Recall that $\mathbf{N}_G$ has n normalized orthogonal eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$ in $\mathbb{R}^n$ associated with n real eigenvalues $1=\lambda_1>1-\alpha\geqslant\lambda_2\geqslant\ldots\geqslant\lambda_n\geqslant\alpha-1$ respectively, and $\mathbf{v}_1\parallel\mathbf{1}\mathbf{D}^{1/2}$. Then $\frac{\mathbf{N}_G+\mathbf{I}}{2}\otimes\mathbf{I}$ has n^2 normalized orthogonal eigenvectors $\mathbf{v}_i\otimes\mathbf{v}_j$ associated with eigenvalues $\frac{\lambda_i+1}{2}$, $i,j=1,\ldots,n$. Write $\mathbf{u}_2=\mathbf{w}_1+\mathbf{w}_2$ such that $\mathbf{w}_1\in V\triangleq\operatorname{span}(\mathbf{v}_1\otimes\mathbf{v}_1,\ldots,\mathbf{v}_1\otimes\mathbf{v}_n)$ and $\mathbf{w}_2$ is in the orthogonal complement of V. Then $\mathbf{w}_1=(\mathbf{1}\mathbf{D}^{1/2})\otimes\mathbf{w}'$ for some $\mathbf{w}'\in\mathbb{R}^n$. Then

$$\begin{aligned}\|\mathbf{w}_1\|_2^2=\langle\mathbf{w}_1,\mathbf{w}_1\rangle=\langle\mathbf{u}_2,\mathbf{w}_1\rangle&=\left\langle(\mathbf{u}\mathbf{Q}^k)^{\perp}(\mathbf{D}\otimes\mathbf{D})^{-1/2},(\mathbf{1}\mathbf{D}^{1/2})\otimes\mathbf{w}'\right\rangle\\ &=\left\langle(\mathbf{u}\mathbf{Q}^k)^{\perp},\mathbf{1}\otimes(\mathbf{w}'\mathbf{D}^{-1/2})\right\rangle\\ &=\left\langle\mathcal{T}_1((\mathbf{u}\mathbf{Q}^k)^{\perp}),\mathbf{w}'\mathbf{D}^{-1/2}\right\rangle\\ &\leqslant\left\|\mathcal{T}_1((\mathbf{u}\mathbf{Q}^k)^{\perp})\right\|_2\left\|\mathbf{w}'\mathbf{D}^{-1/2}\right\|_2\\ &\leqslant(1-\alpha/2)^k\beta^{1/2}\left\|\mathbf{w}'\mathbf{D}^{-1/2}\right\|_2\\ &\leqslant(1-\alpha/2)^k\beta^{1/2}\delta^{-1/2}\|\mathbf{w}'\|_2\\ &\leqslant(1-\alpha/2)^k\beta^{1/2}\delta^{-1}\left\|(\mathbf{1}\mathbf{D}^{1/2})\otimes\mathbf{w}'\right\|_2\\ &=(1-\alpha/2)^k\beta^{1/2}\delta^{-1}\|\mathbf{w}_1\|_2\end{aligned}$$

where the second inequality follows from Lemma 3.19. So $\|\mathbf{w}_1\|_2\leqslant(1-\alpha/2)^k\beta^{1/2}\delta^{-1}$. The matrix $\left(\frac{\mathbf{N}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}$ fixes $\mathbf{w}_1$ and reduces the ℓ_2-norm of $\mathbf{w}_2$ to at most $(1-\alpha/2)^{k'}\|\mathbf{w}_2\|_2\leqslant$

$(1-\alpha/2)^{k'}\|\mathbf{u}_2\|_2$, since $\mathbf{w}_2$ is in the span of eigenvectors $\mathbf{v}_i \otimes \mathbf{v}_j$, $i \neq 1$, that have eigenvalues $\frac{\lambda_i+1}{2} \leqslant 1-\alpha/2$. Then

$$\begin{aligned}
\|\mathbf{u}\mathbf{Q}^k\mathbf{M} - \boldsymbol{\pi}\otimes\boldsymbol{\pi}\|_2 &= \left\|\mathbf{u}_2\left(\frac{\mathbf{N}_G+\mathbf{I}}{2}\otimes\mathbf{I}\right)^{k'}(\mathbf{D}\otimes\mathbf{D})^{1/2}\right\|_2 \\
&\leqslant \left(\|\mathbf{w}_1\|_2 + (1-\alpha/2)^{k'}\|\mathbf{u}_2\|_2\right)\left\|(\mathbf{D}\otimes\mathbf{D})^{1/2}\right\|_2 \\
&\leqslant \|\mathbf{w}_1\|_2\cdot\Delta + (1-\alpha/2)^{k'}\beta\left\|\mathbf{u}_2(\mathbf{D}\otimes\mathbf{D})^{1/2}\right\|_2 \\
&\leqslant (1-\alpha/2)^k\beta^{3/2} + (1-\alpha/2)^{k'}\beta\left\|(\mathbf{u}\mathbf{Q}^k)^\perp\right\|_2 \\
&\leqslant (1-\alpha/2)^k\beta^{3/2} + (1-\alpha/2)^{k'}\beta\cdot(8\alpha^{-1}\beta^7 n^{-3/2} + (1-\alpha/8)^k\beta^2) \\
&= (1-\alpha/2)^k\beta^{3/2} + (1-\alpha/2)^{k'}\cdot 8\alpha^{-1}\beta^8 n^{-3/2} + (1-\alpha/2)^{k'}(1-\alpha/8)^k\beta^3
\end{aligned}$$

where we use the fact that $\left\|(\mathbf{u}\mathbf{Q}^k)^\perp\right\|_2 = \|\mathbf{u}\mathbf{Q}^k - \boldsymbol{\pi}\otimes\boldsymbol{\pi}\|_2 \leqslant 8\alpha^{-1}\beta^7 n^{-3/2} + (1-\alpha/8)^k\beta^2$ by Lemma 3.20. ■